1-1-2012

Informative Bayesian Modeling With Applications to Media Data

Pengyuan Wang
University of Pennsylvania, deezephel@gmail.com

Informative Bayesian Modeling With Applications to Media Data

Abstract
This dissertation consists of three main parts. Each part develops an application or methodology within the Bayesian framework. The first is a study of multi-channel media consumption patterns for US audiences during the 2010 FIFA World Cup using a Bayesian data fusion strategy. We utilize the aggregated television ratings in the estimation, to incorporate additional data that is on a different scale than the individual-level on alternative media platforms. The second study proposes an information integration method, called the information reweighted prior (IRP) approach, to incorporate external information via prior distributions through reweighting. We demonstrate the effectiveness of IRP with both simulated and real panel choice datasets, and show that `sensible' external information, even if with considerable uncertainty, can improve inferences for quantities of interest. The third study proposes a rank enhanced likelihood (REL) approach to utilize ranking information via re-construction of the likelihood. We demonstrate the effectiveness of REL with simulated datasets, and show that utilizing REL can also improve posterior inferences.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Statistics

First Advisor
Eric T. Bradlow

Second Advisor
Edward I. George

Subject Categories
Advertising and Promotion Management | Marketing | Statistics and Probability

INFORMATIVE BAYESIAN MODELING WITH APPLICATIONS TO MEDIA DATA

Pengyuan Wang

A DISSERTATION

in

Statistics

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

Eric T. Bradlow

Professor of Marketing, Statistics, and Education

Co-Supervisor

Edward I. George

Professor of Statistics

Graduate Group Chairperson

Eric T. Bradlow, Professor of Marketing, Statistics, and Education

Dissertation Committee

Dylan Small, Associate Professor of Statistics

Edward I. George, Professor of Statistics

Eric T. Bradlow, Professor of Marketing, Statistics, and Education

INFORMATIVE BAYESIAN MODELING WITH APPLICATIONS TO MEDIA DATA

ACKNOWLEDGEMENT

I would like to thank my advisors Eric Bradlow and Edward George for their support and discussions, which contribute fundamentally to my research and understanding of the nature of statistics. I thank Elea Feit, Eric Bradlow, Peter Fader for their work on our ESPN media usage study and their help about marketing concepts and the related applications of statistics. I thank Dylan Small and Mikhail Traskin for their patient and hands-on guidance when I wrote my very first academic paper. I thank all the faculties in the Statistics Department for the great curricula and inspiration.

The students here are super, and I thank all of them, including the ones who have already graduated, for the discussions and fun.

ABSTRACT

INFORMATIVE BAYESIAN MODELING WITH APPLICATIONS TO MEDIA DATA

Pengyuan Wang

Eric T. Bradlow

Edward I. George

This dissertation consists of three main parts. Each part develops an application or methodology within the Bayesian framework. The first is a study of multi-channel media consumption patterns for US audiences during the 2010 FIFA World Cup using a Bayesian data fusion strategy. We utilize the aggregated television ratings in the estimation, to incorporate additional data that is on a different scale than the individual-level on alternative media platforms. The second study proposes an information integration method, called the information reweighted prior (IRP) approach, to incorporate external information via prior distributions through reweighting. We demonstrate the effectiveness of IRP with both simulated and real panel choice datasets, and show that 'sensible' external information, even if with considerable uncertainty, can improve inferences for quantities of interest. The third study proposes a rank enhanced likelihood (REL) approach to utilize ranking information via re-construction of the likelihood. We demonstrate the effectiveness of REL with simulated datasets, and show that utilizing REL can also improve posterior inferences.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : Introduction

This dissertation consists of three chapters. The first is a study of multi-channel media consumption patterns for US audiences during the 2010 FIFA World Cup using a Bayesian data fusion strategy. We utilize aggregated television ratings in the estimation, to incorporate additional information that is on a different scale than the individual-level that is available for alternative media platforms. The second study proposes an information reweighted prior (IRP) approach to incorporate external information via prior distributions through reweighting. The third study proposes a rank enhanced likelihood (REL) approach to utilize ranking information via re-construction of the likelihood. The three works are all around informative Bayesian modeling. [1]

A basic motivation for the whole dissertation is that, in parametric Bayesian inference, the likelihood is usually selected to fit the data overall. While the model may perform well globally, it may not be able to capture a specific aspect of the data, for example orderings, individual-level predictions, and other 'localized' inferences. We found two general ways to improve the inference of the specific aspect that we are interested in. One is that related external information may be available from various resources, such as surveys, expert knowledge prior studies, etc.. Incorporating the external information may improve the inference of the aspect of interest, but is inconsistent with typically used non-informative priors and hyper-priors. Also, in current marketing research, information arrives at a very rapid pace, and hence methods in marketing that allow for coherent, sequential and fast information integration (updating of beliefs) are needed. IRP provides a new approach to information integration in a model-based setting. In particular, we adapt existing methods that have become popular in marketing, i.e. when a Bayesian model has been fit using Markov Chain Monte Carlo Methods. Specifically, the IRP approach is a sample reweighting approach for sequential information updating which has no restrictions on the likelihood,

[1] The first study is joint work with Elea McDonnell Feit, Eric Bradlow and Peter Fader. The second and third studies are joint work with Edward George and Eric Bradlow.

prior distributions, or data structure; hence a general purpose tool.

For the IRP approach, information resources could be previous or external research (Lybbert et al. 2007, Higgins and Whitehead 1996), experts (Sandor and Wedel 2001), theories (Montgomery and Rossi 1999), or external datasets (Lind and Kivisto-Rahnasto 2008, Lenk and Rao 1990, Putler et al. 1996, Wedel and Pieters 2000, Hofstede et al. 2002), or even the dataset under research (Chapter 4). It may also be considered as a method for information integration (or meta-analysis, Sutton and Abrams (2001), Trikalinos et al. (2008), and an application of meta-analysis with informative priors as in Higgins and Whitehead 1996).

Currently, existing methodologies to incorporate additional information are usually based on a specific kind of information. For example, data augmentation and Bayesian missing data approaches, (Chen and Yang (2007), Musalem et al. (2008)) are usually applied to additional data on aggregated level when analyzing individual-level panel datasets. This method is utilized in the application in Chapter 2. In this chapter, we investigate multi-channel media consumption patterns for US viewers during the 2010 FIFA World Cup, the most watched sporting event in the world. Accordingly, this was the first time a major media company provided live coverage of a tournament through four distinct channels: TV, web, online streaming video and mobile, thus presenting a unique opportunity to study multi-channel media consumption. As media channels proliferate, a key question facing media providers is whether the investment in media coverage on an additional channel will increase total reach and frequency across all channels (and therefore advertising revenue). In this chapter, we develop a hierarchial Bayes multivariate logit model to explore the relationships between media platforms and to project total reach had the tournament not been covered on all platforms. Through a dual covariance structure across platforms, the model captures the correlations in usage across platforms both over the entire course of the tournament and on individual days. The data set we present in this chapter is typical of those collected by today's multi-platform content providers, consisting of individual-level records of daily usage for the digital platforms and aggregate ratings data for television. Using a Bayesian data fusion strategy, we show how the aggregated ratings can be incorporated in the estimation of the individual-level multivariate logit model. The modeling approach we propose can be used by other content providers to understand the changing relationships among platforms as new technologies emerge and media consumption evolves, even when the data available for analysis is measured at different levels of aggregation across channels.

However, such methods are usually application-specific, and not generalizable to a unified system. Chapter 3 proposes an information reweighted prior (IRP) approach to incorporate external information via prior distributions through reweighting. It is a unified approach which assumes no restriction on the likelihood, prior distributions, or data structure. When the external information comes from previous research, the IRP approach can be viewed as a new Bayesian meta-analysis procedure. We demonstrate the effectiveness of IRP with simulated panel choice datasets, and show that 'sensible' external information, even if with considerable uncertainty, can improve posterior estimation for the inferential goal, in comparison to standard Bayesian analyses. As a real-world application, we apply the IRP approach to a unique online advertising dataset with external information summarized from previous online advertising literature (both academic and practitioner) in one case, and from out-of-sample summaries of the dataset in another.

The other general direction is to further explore the data to better utilize the information of the aspect that we are interested in. Again in parametric Bayesian inference, the likelihood is usually selected to fit the data overall. While the model may perform well globally, it may not be able to capture a specific aspect of the data, for example rankings and orderings, a common managerial inferential problem. Hence in Chapter 4, we propose a rank enhanced likelihood (REL) approach to utilize ranking information via re-construction of the likelihood. The framework is within a Bayesian framework and inferences from the model are obtained from posterior samples using Markov Chain Monte Carlo techniques. We demonstrate the effectiveness of REL with simulated datasets, and show that utilizing REL can improve the inference of the related aspect in comparison to standard Bayesian analyses. As a real-world application, we apply the REL approach to a sales scanner dataset, a common setting in which ranking of brands is desired. Even though Chapter 4 is only describing rank enhanced likelihood, it could be readily generalized to other aspect of interest. Also, as a comparison to the way where one incoporates information of the aspect of interest via priors, as in the IRP method, the enhanced likelihood method enables the aspect of interest to play a key role in the setup of likelihood, and hence more directly on inference.

The rest of the thesis is organized as follows. In Chapter 2, we conduct the data-fusion application which incorporates additional information based on aggregate TV ratings. In Chapter 3, we propose the IRP method and compare the performance of IRP and benchmark Bayesian analyses via simulations and perform a real-world application to an online advertising dataset from Organic Inc.. In Chapter 4, we propose the rank enhanced likelihood method with simulated examples and a study on and ERIM grocery shopping dataset. In Chapter 5, we conclude with a discussion for future research.

CHAPTER 2 : Modeling Multi-Platform Media Consumption for the FIFA World Cup Utilizing Aggravate and Disaggregate Information

2.1. Introduction

When fans wanted to know what happened on a particular day of the 1990 FIFA World Cup Tournament, they had few choices: they could see a final score on the evening news and then read about it in a newspaper the following day. But in 2010, in sharp contrast, World Cup fans could choose among numerous media platforms to follow every game in real time: they could follow constantly updated coverage on a traditional website or on their mobile phone, they could watch every game live on television, or they could watch via online streaming video, either live or recorded. By the time the 2014 World Cup takes place, fans will likely have even more ways to access tournament coverage, perhaps including interactive television, mobile apps, or any number of other media platforms that have not even been invented yet. Clearly, from a business perspective, the multitude of media delivery platforms provides both opportunities for greater media exposure and higher advertising reach (the "currency" of the media business), but also greater challenges for media companies that are faced with deciding whether or not to invest in developing content for (and steering audiences to) each new platform. For these companies, understanding media consumption across platforms is critical to driving viewership and advertising revenue. It is from this perspective that we motivate this study.

2.1.1. Media Planning Challenges and Opportunities

As platforms have proliferated and opportunities to reach media viewers have increased, prices for advertising are at risk of falling and so profitably operating a media platform has become more complex. In order for a new media platform to be profitable, it must attract enough viewers to recoup the non-trivial costs of creating content specifically for that platform. However, it is very difficult to assess the contribution of a new media platform within a multi-platform media system. A new platform may gain reach by cannibalizing viewers that might have used an existing platform, were the new platform not available. And if advertising is sold at a lower price on the new platform, this may actually decrease profits. For instance, some users may quickly glance at scores on a mobile device instead of spending time watching a full game on TV. Or the opposite may occur; a new, more accessible platform may improve users' media experience substantially and increase overall interest in an event such as the World Cup. For example, keeping up with the tournament scores via mobile coverage may spur users to watch more games on TV. Media planners can no longer analyze usage of each platform independently, as was once common practice, and are are in need of tools that can be used to disentangle the contribution of each individual platform as part of the multi-platform system.

The few studies that have explored consumers use of multiple media platforms are generally based on surveys where consumers keep a written diary of their media consumption (e.g., Lin et al. (2010), Pilotta et al. (2004), Pilotta and Schultz (2005)). This reliance on survey-based methods is somewhat puzzling as the same digital technology that is proliferating new media platforms also provides a rich stream of behavioral data that can be used to answer media planners' questions. Unlike broadcast television, newspapers, and radio, most of the emerging media platforms create a record of every user and what he or she viewed. Digital media, including the Internet and mobile devices, provide rich streams of data describing which customers consumed content on which platform in each moment. This data is routinely tracked using tools such as Adobe SiteCatalyst and IBM Coremetrics. As media consumption migrates to these well-measured platforms, companies will regularly have the data necessary to constantly monitor how users are using multiple platforms at the same time. Although our data and case study is limited to a particular type of content at a particular point in time, most major content providers collect data similar to the data we use and the basic modeling approach we propose could be used by other content providers to measure and understand the changing relationships between platforms as media consumption behavior evolves and new platforms are introduced.

We focus here on data that is measured on the server-side (also known as *site-centric* data as in Zheng et al. (2009)), which is essentially a census of all users of a particular platform and is readily available to media companies. The core data in our case study is a random sample of 2,000 ESPN digital media users. For these users we observe daily consumption across three platforms: a traditional website, online streaming video and mobile.

However, like most media companies, ESPN does not have ready access to individual-level behavioral data on television viewing for these 2,000 users. Globally, and specifically in the US where our data was collected, only a small fraction of households have transitioned to measurable television systems, and even if television usage is recorded for individual households by the cable or satellite provider, technology differences between television and the Internet make it difficult to link a particular household's television usage to their digital media usage. This presents a challenge typical for today's media companies: the company wants to understand users' multi-channel media consumption, including television, and it has rich data on individuals' digital media usage, but they are lacking detailed data on television viewing behavior. The only readily available data on television viewing is aggregate ratings data. So, as is typical for most media companies, we have data in a mixed structure: panel data for digital platforms (where individual usage is tracked daily) along with aggregate daily television viewership over the same time period.

2.1.2. Modeling Approach

Our objective is to develop a modeling approach that can be used with this type of readily available mixed aggregate/disaggregate, cross-platform data. Using a Bayesian modeling framework allows us to specify the model as if we had individual-level television consumption data and then, using data augmentation, we integrate the likelihood over all possible values of the individual-level television usage data that are consistent with the observed aggregate consumption data. While this approach to aggregate data as a form of "missing data" is not new (c.f., Chen and Yang (2007), Musalem et al. (2008)), our contribution is in showing the usefulness of this approach for multi-channel media consumption data. This is also, to the best of our knowledge, the first application of such an approach to a mixed aggregate/disaggregate data structure, and certainly is the first application to modeling media consumption.

Our ultimate goal is to understand the correlations in usage among platforms, so that we can identify whether usage of two (or more) platforms is positively or negatively correlated. However that measurement is complicated by the fact that media consumption patterns vary widely among consumers. For example in our data set, many users were inactive on all 38 days we observe, while other users consumed content on nearly all of the days. This, combined with the sparseness of the data for any individual user, motivates our use of a hierarchical Bayesian multivariate logit model featuring a vector of media-platform-specific intercepts for each consumer that incorporates shrinkage. Within this structure, we allow for negative or positive correlations among the platform intercepts for each user, as well as negative or positive correlations among daily error terms for each user. This allows us to distinguish the daily substitution among channels, i.e., people who use one platform on a given day may be less likely to use another platform *on that same day*, versus the long-term positive correlations that are typically found in media consumption behavior, i.e., people who consume more often on one platform are more likely to consume more on other platforms over time. Somewhat jokingly, therefore, this is one of those rare occasions where elements of covariance matrices are truly parameters of interest (not just nuisance parameters to soak up unexplained variation) in order to address the aforementioned substantive issues. For instance, with these parameters that measure the relationships among media channels, content providers can begin to understand whether new channels are detracting from or enhancing consumption on existing channels, and therefore whether and how much to invest in them. We demonstrate, through a synthetic data study reported in the appendix, that the intra-day correlations among platforms (but not the long-term correlations across user intercepts) can be recovered even when usage for one of the platforms is only measured in aggregate.

Our statistical model is general in that it can be applied to any mixed individual-level/aggregate data set covering multi-platform media usage; however in our case study we apply it to data on media consumption from the 2010 FIFA World Cup tournament. This tournament was viewed by more than 100 million Americans and more than 1 billion people globally, making it the most-viewed sporting event in the world. The unparalleled scale of this tournament makes studying it of particular interest to ESPN as an important test case for answering questions about investments in multi-platform coverage. The FIFA tournament structure, however, represents a particular modeling challenge: consumption of World Cup media content is clearly driven by which individual teams play on a given day. For example, the peak daily consumption during the tournament was more than three times the consumption on any day prior to the start of the tournament, yet there are days during the tournament that have consumption near that low pre-tournament baseline. Any reasonable model for this data should therefore control for the content of the tournament. We do this using a set of covariates that are specific to a such an event, including the number of games on a given day, whether it was a weekend or weekday, and whether the US team was playing. [2] By estimating these effects separately for each platform, we gain some additional insight into how the platforms differ in terms of what types of content are most appealing.

2.1.3. Related Literature

The issue of how users interact with multiple platforms has been of interest to media planners particularly as new media platforms have started to proliferate (e.g., Franz (2000)). A key question in the literature is whether consumers use multiple media platforms at the same time.[3] Prior theoretical work has shown that rational media consumers will use multi-

[2] We also explored the possibility of using a shrinkage model to estimate an effect for each team, but this proved difficult with the available data. A total of 32 teams begin the tournament and were assigned into 8 groups of 4 teams each and play a round-robin with the other teams in their group. Based on the results of this group stage, half of the teams (2 from each group) proceed to a knockout-style, one-and-done, elimination tournament. This results in a tournament schedule where half of the teams only play on 3 days, giving us limited information from which to infer attractiveness effects for each of them.

[3] The vocabulary to describe this behavior is still evolving. *Multihoming* is used in the theoretical literature to refer to the situation where consumers use one or more platforms, but not necessarily at the same time (c.f. Parker and Van Alstyne (2005)), while *multiplexing* has been proposed to refer to consumers who use two media platforms *at the same time* (Lin et al. (2010)).

ple platforms under certain circumstances (c.f. Parker and Van Alstyne (2005)). However, most of the empirical work in this area only *suggests* that individual users might be using multiple media platforms. For example Joo and Zhu (2011) find that there is a relationship between the airing of television advertisements and aggregate online search behavior, suggesting that individual users must be using TV and search at the same time. Similarly, aggregate marketing mix models have shown synergy effects among advertising channels (Naik and Raman (2003), Naik and Peters (2009)) and the most likely explanation for this synergy is that individual consumers are viewing content on multiple channels and that the cross-channel repetition is highly effective. So there is great interest among media planners in the issue of whether viewers use multiple media platforms, but there have been few empirical studies that have investigated individual-level multi-channel media consumption directly, partly because of data scarcity. Prior survey-based studies of multi-platform user behavior have also focused on advertisers' questions (e.g., which advertisements should I place where to maximize sales?), while our case study focuses on the media planning problem from the content provider's perspective (e.g., which platforms should I invest in to maximize audience and advertising opportunities?)

While there has been some prior work on multi-channel media consumption, our work is methodologically more akin to empirical work that has developed models for closely related data structures. Our core data structure is one where consumers engage in multiple activities over time and so is quite similar in structure to data on which websites users are visiting over time (Danaher (2007)), which categories consumers are purchasing from over time (e.g., Manchanda et al. (1999)) and which distribution channels a customer is using over time (e.g., Ansari et al. (2008)). However, while all of these papers model an individual's multi-platform usage over time, none has tackled the problem when data on usage of one of the platforms is only available in aggregate. This is somewhat surprising, given how frequently this general structure occurs in practice. For example, the typical multi-channel retailer may have detailed transaction data for their direct channels such as online and catalog, but can only track brick-and-mortar sales at the aggregate level. To resolve this problem we draw on Bayesian missing data approaches applied in other contexts with different types of data structures (Chen and Yang (2007), Musalem et al. (2008)).

The remainder of the chapter proceeds as follows. In the next section, we describe the data set acquired through ESPN on daily media consumption during the 2010 FIFA World Cup. This is followed by a description of our model and a set of posterior predictive checks (Gelman et al. (1996)) that will be used to illustrate the fit of the model to the data. We then report and interpret the estimated parameters, as well as a set of counterfactual analyses predicting ESPN usage had coverage not been provided (fully or in part) on the mobile platform. We conclude with thoughts for future applications of our modeling framework to a more general class of problems, as well as a data "wish list" for future research on multi-platform media consumption.

2.2. ESPN Media Consumption Data for the World Cup

We observe media consumption for a sample of 2,000 ESPN registered users[4] of ESPN's interactive media services for each day from June 4, 2010 to July 11, 2010. The observation window includes the week prior to the start of the World Cup tournament, which began on June 11, 2010 and concluded with the final championship game on July 11, 2010. This project was part of a larger initiative (with many participating media research firms), called ESPN XP (Crupi (2010)), designed to help ESPN and their advertisers better understand cross-platform user behavior.

For each user, we observe a binary vector indicating whether he accessed each of the three digital channels on a given day. For each user on each day, we observe whether he watched or read *soccer-related content* on each of three media platforms: (i) ordinary digital content on the regular magazine-format website (ESPN.com), (ii) live or archived streaming video of full games available on the ESPN3.com platform (a separate website from ESPN.com)[5],

[4]While we recognize that registered users are not a perfectly representative sample of ESPN's population of users, they do represent a group that can be tracked longitudinally over time and because ESPN has invested in getting a large proportion of their users to login, the selection bias is not too severe. It is certainly less severe than the bias in a typical opt-in media consumption panel.

[5]Although both ESPN.com and ESPN3 are both viewed on the Internet, from the point of view of the

and (iii) ESPN Mobile (a site formatted for mobile phones and tablet devices with a smaller screen). Because one of ESPN's objectives was to encourage those who had not previously watched soccer to follow the World Cup, the sample does not condition on the user being a soccer fan or even having consumed soccer content prior to the tournament; thus, it is not surprising that about half of the sampled users, who are primarily based in the US, do not access any soccer-related content during the tournament. Because the focus of this project was to better understand multi-platform behavior, and because many US users did not yet have smart phones or other mobile devices at the time of the tournament, the target population from which we sampled were users who were observed to use mobile services (for any content) sometime in the year prior to the tournament. These users represent what we might consider the "vanguard" of mobile users.

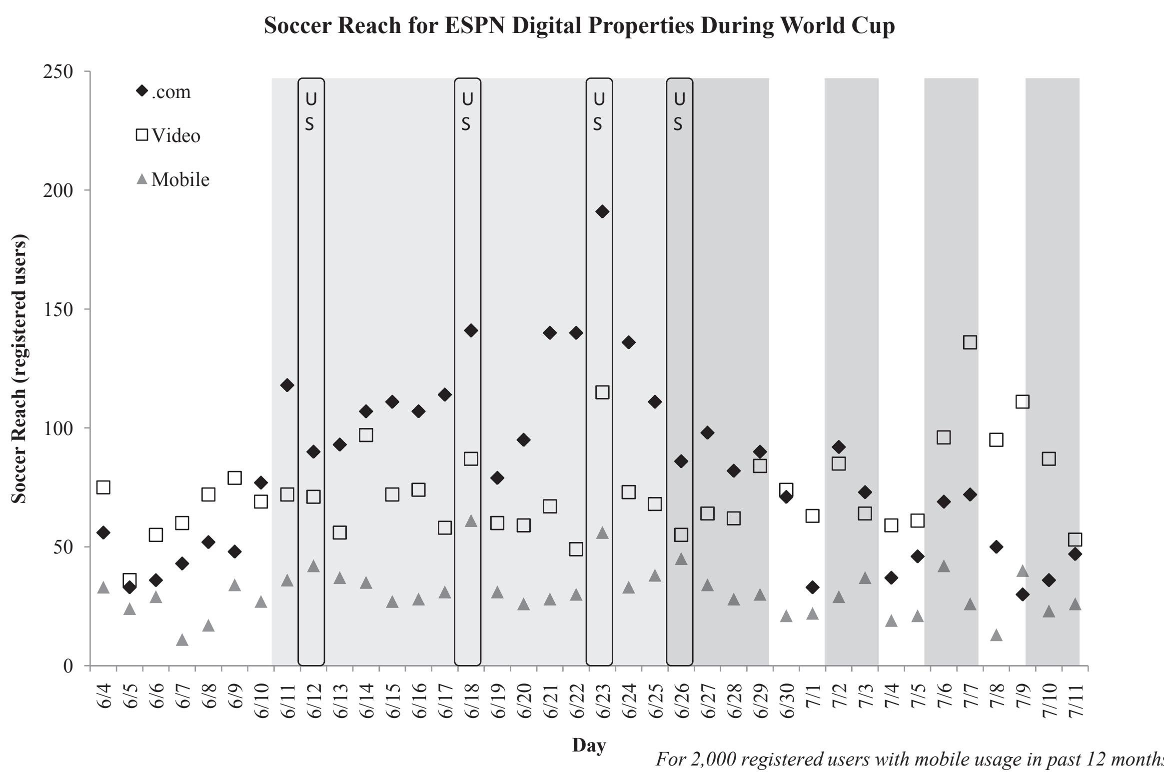
We combine this digital media data with aggregate data on television ratings provided by The Nielsen Company. Specifically, Nielsen provided their estimate of the total fraction of US households that watched any of the televised English-language broadcasts of World Cup games for each day during the tournament.[6] Like all data-fusion problems, merging the ratings data with the digital media data, requires some assumptions about the relationship between the two data sets. Our core assumption is that the fraction of our digital users who watched the TV broadcast of the World cup approximately matches the fraction of total US households who watched on TV as reported by Nielsen. That is, we assume there are no differences in soccer TV viewing habits between the general population and our sample of digital platform users. Since television is the most popular media platform today, we believe this assumption to be reasonable.

user, ESPN3 is a distinct platform with an entirely different interface than ESPN.com, and is only available to users who subscribe to certain cable providers. Unfortunately, from a back-end technology perspective, the video for both sites was hosted on the same server and ESPN co-mingled data on usage of ESPN3 and streaming video embedded in ESPN.com. However, since nearly all the video on ESPN.com was in the form of short clips, with full-game video reserved for ESPN3, we were able to approximately measure ESPN3 usage by only counting the user as having watched ESPN3 if he viewed a video for more than 3 minutes.

[6]We focused our analysis specifically on English speakers because the ESPN.com, ESPN3 and ESPN Mobile audiences are English-language oriented. We excluded from our analysis broadcasts on Spanish-language television and the relatively smaller number of users of ESPN's Spanish-language website, ESPN Reportes.

Summing these observations for each day, we can compute the daily reach for each platform, which is defined as the total number of distinct users who are exposed at least once to a particular medium during a given day. Media providers and advertisers are interested in reach because it represents the total number of people who could be exposed to an advertisement if it was placed on that medium (Rossiter and Danaher (1998)). In Figure 1, we plot the total reach (among the aforementioned 2,000 users) over time for soccer-related content on each of the four platforms.

Figure 1: Daily Soccer Reach for ESPN Media Platforms During the 2010 Fifa World Cup



Days on which tournament games occur are shaded. Days on which the US team was playing are highlighted in boxes.

A number of interesting patterns are readily apparent in Figure 1. First, daily reach for digital platforms tends to decline during the latter part of the tournament. The data also suggest that, for this US-based audience, there is higher reach on all platforms when the

US team is playing. It is not apparent from this plot of the data, but we also find that when teams from countries with strong cultural or geographic ties to the US (e.g. England, Australia or Mexico) play, there is higher average reach as well.

A substantial proportion of users in this data set use multiple platforms to access soccer content, and our ability to forecast and describe these patterns is clearly essential to understanding the relationship between platforms. The first column in Table 2 shows the fraction of users who use each possible combination of the three digital platforms over the course of the entire tournament. Notably, there are a number of users (13.2%) who use both .com and ESPN3, much more than we would expect if usage of those platforms were independent. Relationships between the mobile platform and the other two platforms are difficult to discern without further modeling and analysis. With these descriptions of the

Table 1: Comparison of Predicted and Actual Multi-Platform Usage

	Actual	Predicted	
		Mean	Quantile
No access	**.631**	.605	.66
Only .com	**.043**	.045	.34
Only ESPN3	**.058**	.058	.45
Only Mobile	**.064**	.063	.48
com & ESPN3	**.132**	.125	.75
com & Mobile	**.018**	.021	.27
ESPN3 & Mobile	**.017**	.021	.17
All three	**.057**	.062	.27

data in hand, we now describe the Bayesian hierarchical model we use to analyze this data.

2.3. A Multivariate Hierarchical Bayesian Logit Model for Multi-Platform Usage

We focus first on developing a model for binary outcomes, y_{ikt}, indicating whether a user, $i = 1, ..., N$, accessed content on a given platform $k = 1, ..., K$ on day $t = 1, ..., T$. After laying out the model for complete data, we will discuss modifications for the situation where one or more platforms are observed only in aggregate.

Because there are a large number of users in the sample who do not access soccer content at all during our observation period, we model each user as either completely inactive (zero on all platforms for each and every day) or active with probability p_{active}; a classic "spike-at-zero" mixture model (Morrison and Schmittlein 1988). We will use I_i to indicate (latently) whether user i is active.

Conditional on being active, we model y_{ikt} with a multivariate hierarchical logistic regression given by:

$$y_{ikt}|I_i = 1 \ \sim \ Bernoulli(p_{ikt}) \tag{2.1}$$

$$logit(p_{ikt}) \ = \ \mu_{ik} + x_{kt}\beta_k + e_{ikt} \tag{2.2}$$

where x_{kt} is a vector of covariates describing the content available on platform k on day t, which is multiplied by a vector of platform-specific coefficients β_k. The residual appeal of the platform after controlling for the covariates, μ_{ik}, is user-specific. The vector $\mu_i = (\mu_{i1}, ..., \mu_{iK})$ is assumed to be normally distributed across the population:

$$\mu_i|\mu, \Sigma_\mu \ \sim \ N_K(\mu, \Sigma_\mu), iid \tag{2.3}$$

The mean of μ_i is included to accommodate differences in overall usage rates among active users across the platforms. The matrix Σ_μ captures the covariance among users' propensities to use each of the platforms over the course of the tournament and, as mentioned earlier, is one of the sets of model parameters of most interest as it represents the covariance among baseline usage propensities across platforms *over time*. The error term $e_{it} = (e_{i1t}, ..., e_{iKt})$ from equation (2.2) is also modeled as a multivariate normal:

$$e_{it}|\Sigma_e \ \sim \ N_K(0, \Sigma_e), iid \tag{2.4}$$

allowing for correlations among the customers' propensities to use each of the platforms *on a given day* through Σ_e.[7]

[7] This specification assumes that e_{it} is independent of $e_{i,t+1}$ and one might ask whether this assumption is warranted. That is, are the customers more likely to watch today conditional on having watched yesterday? To confirm this assumption, for our case study, we calculated the lag-one autocorrelation of the residuals

If y_{it} is the vector of platform usage for user i in period t, the full likelihood for the hierarchial model is given by:

$$\begin{aligned} \ell\left(\{y_{it}|I_i=1\}\right) &= \prod_i \left(\prod_t \left[y_{it}|\mu_i,\beta_k,e_{it}\right]\left[e_{it}|\Sigma_e\right]\right)\left[\mu_i|\mu,\Sigma_\mu\right] \\ &= \prod_i \left(\prod_t \left(\prod_k \left(\frac{p_{ikt}}{1+p_{ikt}}\right)^{y_{ikt}}\right) N_K(e_{it}|0,\Sigma_e)\right) N_K(\mu_i|\mu,\Sigma_\mu) \end{aligned} \tag{2.5}$$

In summary, the model we propose for consumers' media consumption across multiple platforms is a hierarchical multi*variate* logit model with a "spike at zero" for a consumer's use of multiple media platforms. Our model extends the basic multivariate logit framework (Glonek and McCullagh (1995)) by decomposing individual-level cross-day platform effects via Σ_μ from within-day cross-platform effects via Σ_e. This structure allows us to determine the short- and long-term correlations among media platforms, which, as we explained in the introduction, is critical to media planning. The hierarchical Bayes framework (Gelman et al. (2003)) also allows us to determine which users are most likely to watch each platform through the estimates of μ_{ik}. We tested a similar hierarchical multivariate probit specification (Chib and Greenberg (1998); Rossi et al. (2005)) and found it to have similar fit. While the specification of the covariates (x_{kt}) is motivated by our case study, the general model form could be applied to many other multi-platform digital content providers.

2.3.1. Covariates to Control for Tournament Content

The covariates we specify in x_{kt} were selected to account for the relationship between the tournament content and the viewership. Without controlling for spikes in reach that are driven by tournament content (as seen in Figure 1), it would be difficult to interpret the correlations among platforms that are our primary interest.

It is commonly understood that television reach is higher on weekends and online reach is higher on weekdays, reflecting the fact that televisions are generally more accessible on weekends and viewers turn to online coverage when they are at work. This motivated the

for each platform and each customer $i = 1, 2, ..., 2000$. None of the residual sequences showed significant autocorrelation.

inclusion of a dummy variable indicating whether the day was a weekend. This allows us to confirm these previous findings for television and online, as to well as determine whether the new mobile platform displays a similar or different pattern as online or television.

Turning to the tournament itself, there is little prior literature on what might drive viewership in this setting. We assume that the teams that play on a given day have a substantial impact on reach and therefore introduce a set of covariates to control for such effects including: (1) a dummy for whether the US team played, (2) a dummy for whether any of the three top teams (Brazil, Spain and Netherlands according to the May 2010 FIFA scores) played, and (3) a dummy indicating whether any of the three teams with strong cultural connections to the US (England, Mexico, and Australia) played. Assuming that more games and more "critical" games might drive additional reach, we also include variables for the total number of games on a given day and the number of teams who would be eliminated from the tournament on that day if they failed to win, i.e., the number of teams who must "win or go home".

While we will present the parameter estimates for these effects and interpret them, we caution that the estimates are subject to the usual potential biases due to collinearity, missing variables, missing interactions and other misspecifications. However, like any other regression analysis with non-experimental data, we find that the parameter estimates for these effects provide some actionable insights for a media company that wants to understand the "drivers" of multi-channel media viewership. For instance, as we will discuss when we report the parameter estimates, understanding whether mobile usage is higher or lower on weekends (controlling for other aspects of the content) can provide insight into whether mobile coverage on weekends should be emphasized or deemphasized.

2.3.2. Data Fusion Approach

As we described in the introduction, multi-platform content providers often don't have the data required to estimate the model described above for the full set of platforms of interest.

Typically, they have detailed panel data suitable for estimating the above model for the newer digital platforms, but usage data for traditional media channels (radio, television, print) is typically only available in aggregate. In this section, we describe our use of a Bayesian data fusion approach for combining aggregate and disaggregate data (Chen and Yang (2007), Musalem et al. (2008)).

Although the method can be applied when more than one channel is observed in aggregate, for simplicity of exposition, we assume there is only one platform observed in aggregate and that this is the K^{th} platform. For this platform, we do not observe y_{iK}, but instead for each time period, we observe $Y_t = \sum_i y_{iKt}$. The likelihood for the observed disaggregate and aggregate data can be obtained by integrating the likelihood of the model over all possible values of $\{y_{iKt}\}$ that meet this constraint as follows:

$$\ell\left(Y_t\ \right) = \int_{\{y_{iKt}\} s.t Y_t = \sum_i y_{ikt}} \left(\prod_i \left(\prod_t \left[y_{it}|\mu_i, \beta_k, e_{it}\right]\left[e_{it}|\Sigma_e\right] \right) \left[\mu_i|\mu, \Sigma_\mu\right] \right) \tag{2.6}$$

where the integral is taken over all possible values of the set $\{y_{iKt}\}$ that meet the sum constraint implied by the observed aggregate viewership. This model can be estimated in the Bayesian framework through data augmentation of y_{iKt} using standard MCMC methods (Tanner and Wong (1987)).

We note that when the K^{th} platform is only observed in aggregate, the last row and column of Σ_μ corresponding to the covariance between the propensity to use the aggregate platform and the other platforms *on a given day* is not identified. Intuitively, we never observe which individual users using are using the platform on a given day, so it is impossible to estimate covariance between the aggregate platforms and the others. However, as we demonstrate in a simulation study reported in the appendix, the covariance in Σ_e is identified through the repeated measures over time. Consequently, we fix the elements of Σ_e associated with the K^{th} platform to zero. Note that this model still allows for an IID Bernoulli error for daily television usage, through equation (2.1).

We obtain posterior samples for the population-level parameters (μ, β_k, Σ_μ and Σ_e) as well as the "missing" sets of individual-level viewership $\{y_{iKt}\}$ using an MCMC sampler implemented in WinBUGS (Speigelhalter et al. (1999)). We use diffuse proper priors as described in the appendix. Posterior inferences are based on 50,000 draws from the posterior after convergence where the first 50,000 draws were discarded based on trace plots and Gelman-Rubin diagnostics against a second chain run from an independent starting point (Gelman and Rubin (1992)). All code and data are available from the authors upon request. In the appendix, we also report the results of a synthetic data study demonstrating that the parameters of the population-level parameters can all be reasonably recovered (i.e. the posterior covers the value of the population-level parameter used to generate synthetic data), even when only Y_t is observed. This suggests that the parameters we estimate, in particular the correlations in individuals' long-term propensity to use the various platforms in Σ_e, are identified by the data/likelihood combination.

2.4. Model Assessment

Our assessment of model fit focuses on a series of posterior predictive checks at varying levels of disaggregation (Gelman et al. (1996)) that serve to evaluate the model's ability to fit features of the data that are important to media planners. Following the usual procedure for computing posterior predictions, we generated 100 posterior predictive data sets using 100 sets of parameters randomly sampled from the posterior draws obtained from the MCMC sampler. (See the appendix for more detail.) We then compared the posterior distributions for these statistics to actuals computed from the data and report the quantile of the observed value within the posterior predictive distribution, to assess the ability of the model to correctly recover these key statistics.

Fit of Multi-Platform Usage Patterns. By comparing the posterior predictions for the percentage of users who use each combination of platforms, we can assess whether the model is picking up the appropriate covariation, i.e., cross-platform usage – a central question of interest. The last two columns of Table 2 report the ability of the model to predict the 2^3

contingency table of aggregate usage over the course of the tournament for each combination of the digital platforms. (Since we do not observe TV usage at the individual level, and can not compute the actual cross-platform contingencies for TV, we do not include TV in this posterior predictive check.) As can be seen from the data, the fit is quite good with the true values all falling within the .15 and .75 quantiles of the estimated posterior, suggesting that the dual covariance structure adequately captures the covariances between platforms that are observed in the data.

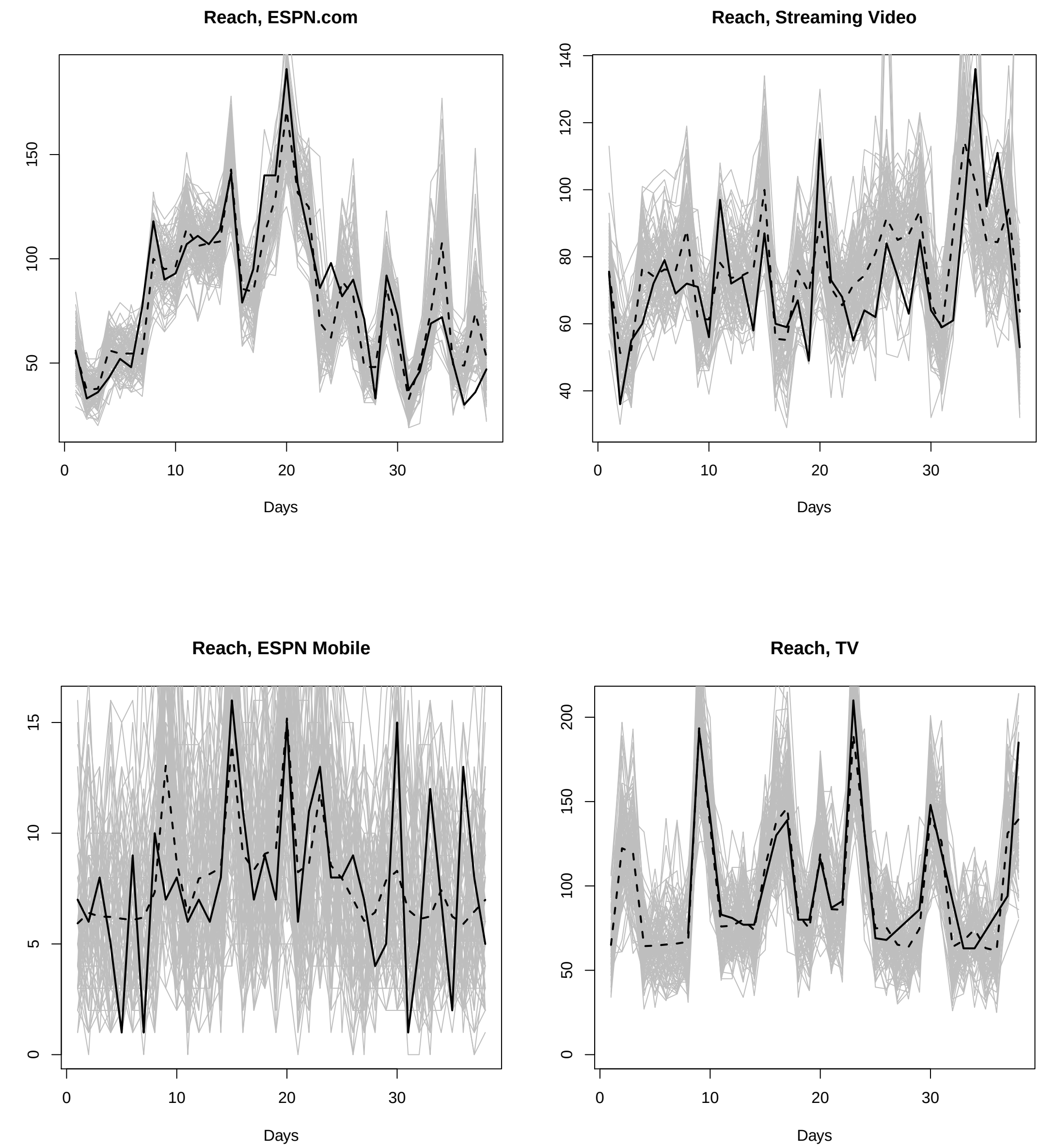
Table 2: Comparison of Predicted and Actual Multi-Platform Usage

	Actual	Predicted	
		Mean	Quantile
No access	**.631**	.605	.66
Only .com	**.043**	.045	.34
Only ESPN3	**.058**	.058	.45
Only Mobile	**.064**	.063	.48
com & ESPN3	**.132**	.125	.75
com & Mobile	**.018**	.021	.27
ESPN3 & Mobile	**.017**	.021	.17
All three	**.057**	.062	.27

Tracking Plots for Daily Reach. Figure 2 plots the posterior mean prediction for daily reach (solid line) compared to the actual daily reach (dotted line) for all four platforms. We also show posterior uncertainty by plotting the daily reach predictions for 100 posterior draws with grey lines. These tracking plots show an excellent fit between the model and the data: the overall mean absolute error between the predicted daily reach (posterior mean) and the actual reach is quite low: .78% for .com, .71% ESPN3, and 1.82% for mobile, suggesting that the model adequately picks up the major features of the total daily reach in the data, including the day-to-day variation.

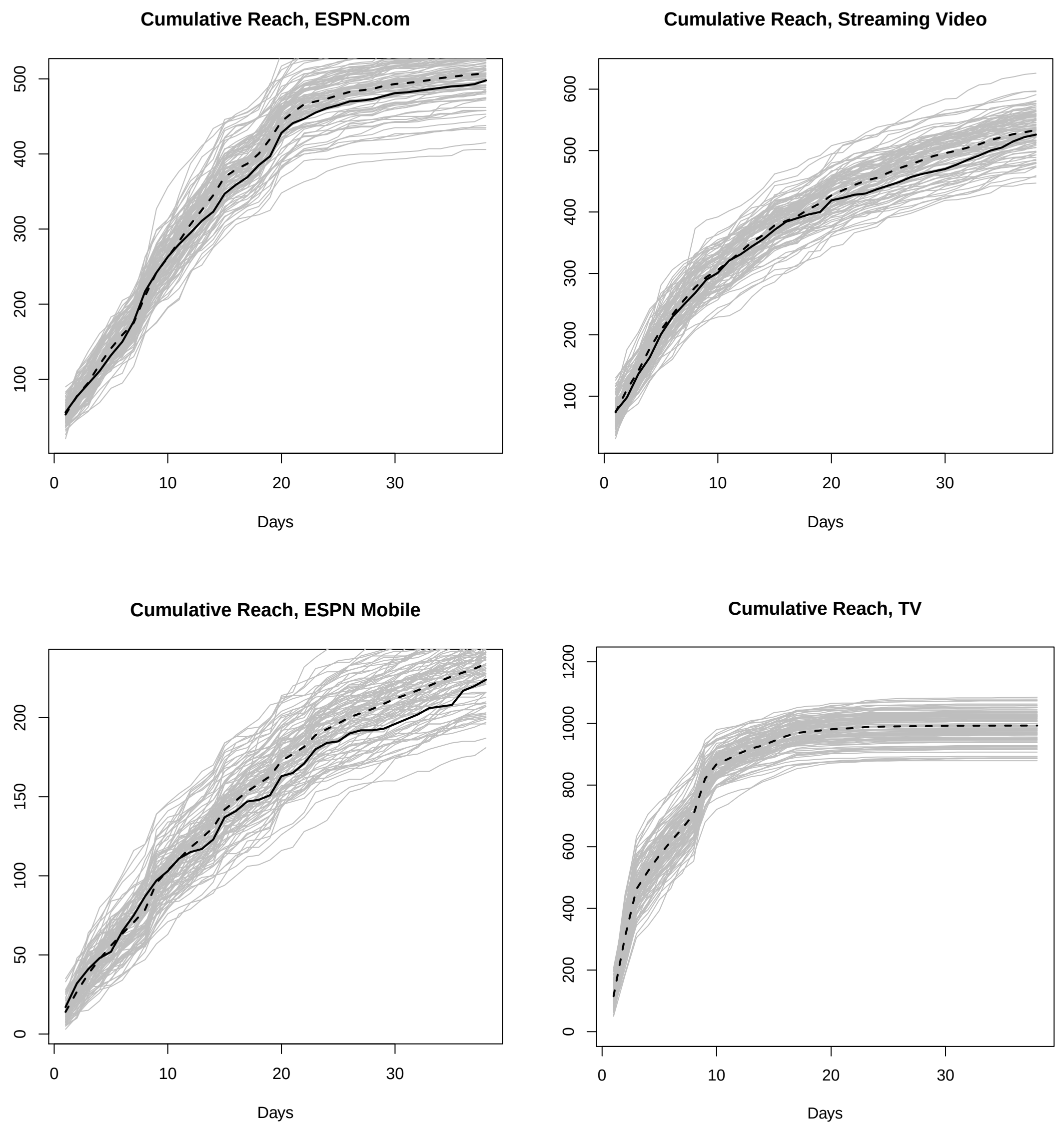
In Figure 3 we show that the model also does an excellent job at predicting cumulative reach for the digital platforms. Cumulative reach, which is defined as the total number of unique users who have viewed content on a particular platform through to a specific day, is frequently used by media planners to understand how many viewers could be reached over the course of an entire media event such as the World Cup. Note that we can not compute the actual cumulative reach for TV, since we do not observe which users are using on each day can not compute the actual; however, we report the model prediction for completeness. The model suggests that cumulative reach for TV levels out about half-way through the tournament, indicating that most of those who will watch a game on TV will watch their first game fairly early in the tournament.

Figure 2: Tracking Plots of Daily Reach for Each Platform



We draw the mean of the generated statistics with a dashed line and compare to the actual statistic (computed directly from the sample) drawn with a solid line.To show the forecast uncertainty, we also draw the prediction for each of 100 random draws from the posterior with grey lines.

Figure 3: Tracking Plots of Cumulative Reach for Each Platform



We draw the mean of the generated statistics with a dashed line and compare to the actual statistic (computed directly from the sample) drawn with a solid line.To show the forecast uncertainty, we also draw the prediction for each of 100 random draws from the posterior with grey lines.

On the TV platform only the aggregate daily usage on game days was observed, and the individual-level usage was not available. Hence when we draw the posterior predictive plots, we can compute the true daily reach, but not the true cumulative reach or true cumulative frequency.

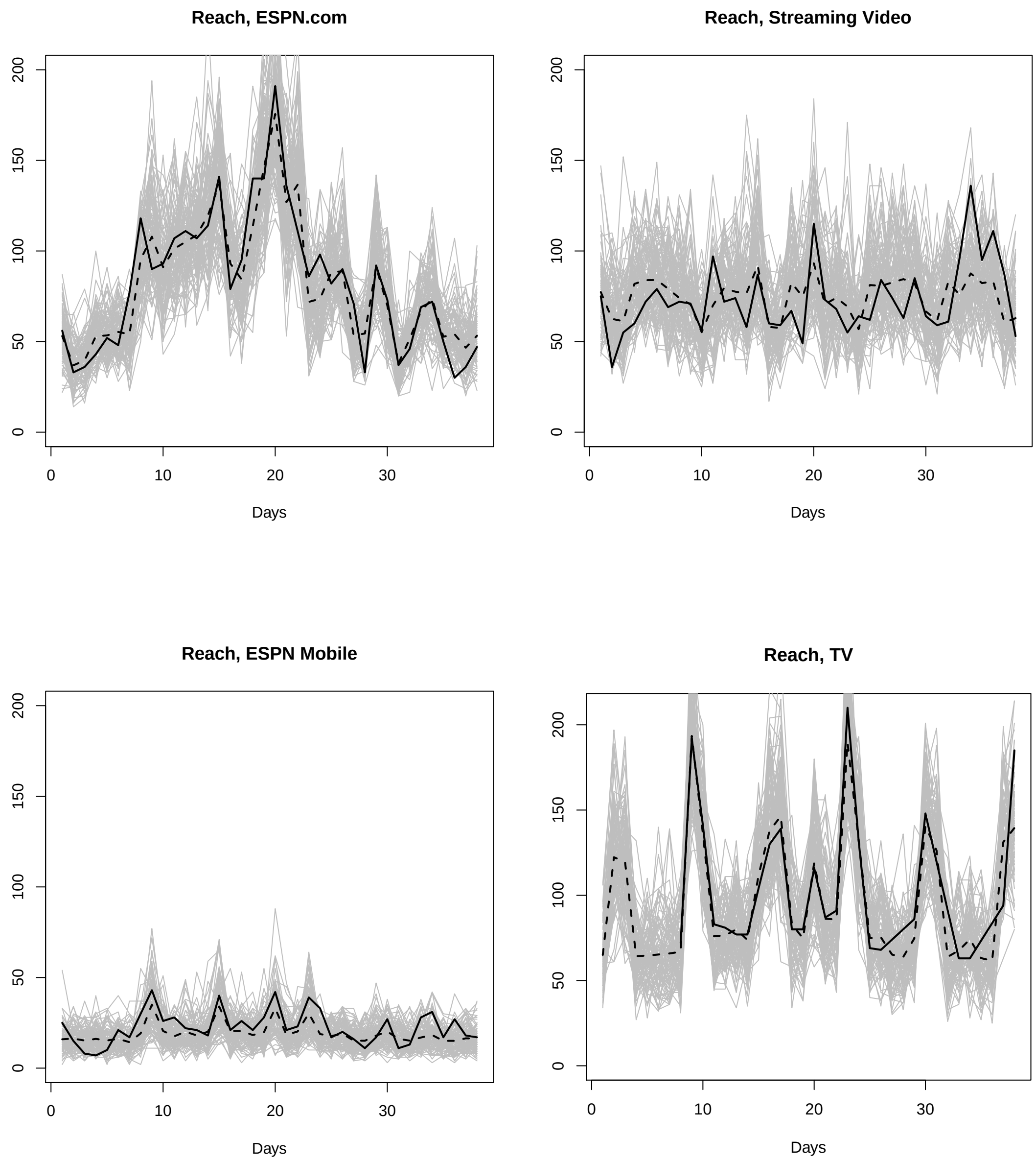
To investigate how well the model fits individuals' media usage patterns, we computed the hit rate for individuals. (We define the hit rate for an individual as the proportion of days the model correctly predicts his usage incidence.) As we see in Table 3, the model does much better than chance at predicting individual's usage with an average hit rates of .922, .929 and .990 for all three digital platforms. (We don't report individual hit rates for TV, since we don't observe individual-level usage on TV.) Table 3 also reports the percentage of users who have an average hit rate over the 38 days that is greater than .5 and .9.

Table 3: Hit Rates for Individual Users

Platform	Average Hit Rate	% of Users with Hit Rate Better Than .5	% of Users with Hit Rate Better Than .95
.com	.922	96.0%	67.7%
ESPN3	.929	98.0%	64.1%
mobile	.990	99.9%	92.8%

To show how the model fits to a new set of users (i.e., a holdout validation), Figure 4 shows the tracking plot of daily reach for a random sample of 2,000 *different* ESPN users who were not used in estimation. As can be seen from the figure, the fit is reasonably comparable to the fit to the estimation data, suggesting no overfitting for this sample of users. (Average Hit Rate = .914, .925 and .979 for the .com, ESPN3 and mobile platforms respectively).

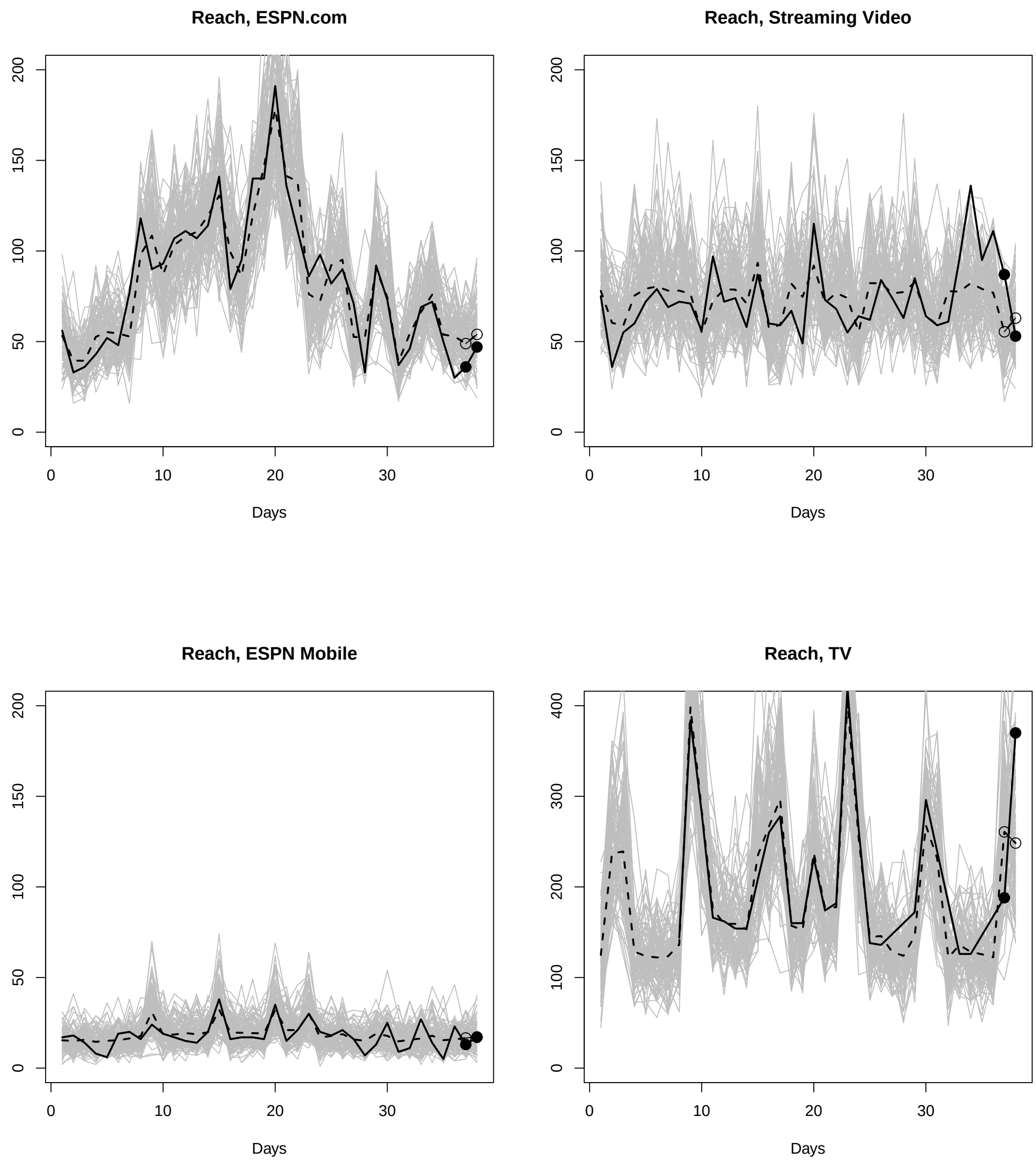
Figure 4: Tracking Plots of Daily Reach for Each Platform for Another Set of Customers



We draw the mean of the generated statistics with a dashed line and compare to the actual statistic (computed directly from the sample) drawn with a solid line.To show the forecast uncertainty, we also draw the prediction for each of 100 random draws from the posterior with grey lines. The figure on the TV platform is the same as the one in Figure 2, since individual level observations on the TV platform are not available.

To show how the model fits to a out-of-sample time period, we fit the model using the observed data before the semi-final and final game days, and then predict the media usage in these two days. Figure 5 shows the tracking plot of daily reach. The fit is reasonably comparable to the fit to the estimation data, suggesting no overfitting for this time period. The fit is slightly worse on the mobile platform. On this platform the usage on days before the final stage is rather flat and begin to surge when entering the final stage, and hence the fitted parameters cannot precisely describe the pattern of the validation data. However, the confirmation of out-of-sample users (Figure 4) and out-of-sample time periods (Figure 5) show confidence about the performance of the model.

Figure 5: Tracking Plots of Daily Reach for Each Platform for Hold-out Time Period



We draw the mean of the generated statistics with a dashed line and compare to the actual statistic (computed directly from the sample) drawn with a solid line.To show the forecast uncertainty, we also draw the prediction for each of 100 random draws from the posterior with grey lines. The semi-final and final game days' observed usage are indicated by black solid points respectively and the mean of the generated statistics are indicated by black circles.

In summary, we find that the model we propose fits the data well, capturing key aspects of the data: aggregate daily and cumulative reach, the pattern of co-usage between platforms and individuals' daily usage. In the appendix, we also report several additional posterior predictive checks, showing that the model does a reasonable job at recovering the heterogeneity between users (i.e., how many light and heavy users there are) and the patterns of usage over time (i.e., how many users "jump in" and "drop out" during the tournament). With this assurance that the model accurately reflects these key aspects of the data, we next turn to interesting estimated parameters of the model.

2.5. Parameter Estimation and Implications for Media Planning

We will discuss, in turn, the parameters that describe the attractiveness of each platform, the heterogeneity around their means, the correlations of platform preferences over time (Σ_μ) and in intra-day platform usage (Σ_e). Following that, we present the platform-specific effects for the tournament characteristics, x_{kt}.

2.5.1. Platform Intercepts and User Heterogeneity

The posterior mean value for p_{active} is .498, indicating that of our sample, 49.8% of the registered users are "active," i.e., they have some predicted probability of accessing soccer content, during the World Cup tournament. Thus, almost half the sample is estimated to be in a spike-at-zero, for each day and each platform, suggesting the necessity for this part of the model. Note, this is consistent with the raw data, in which 61.3% of our sample of 2000 users have no observed usage on the digital platforms.

The general attractiveness to users of the ESPN.com, ESPN3, mobile, and TV platforms are reflected by the parameter vector $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)$. As we can see in Table 4, the intercepts for each platform are in the range of -7 to -2 on the logit scale, suggesting baseline (i.e., on non-tournament days) usage of soccer content on the digital platforms is close to zero. This is consistent with the data where, for those users who visit at least once, we see heterogeneous marginal proportions ranging from a low of 2.6% to a high of 100%. Unsurprisingly, we find that TV is the most popular platform at baseline ($\mu_4 = -2.06$).

Managerially, it is important to keep in mind that these estimates are based on a sample of users who have been known to use the mobile platform (possibly for non-soccer-related content). Although we have earlier characterized these user as the vanguard of mobile device users, it is interesting that they do not strongly prefer the mobile platform over the others. In fact, the mobile platform has the lowest estimated population mean for the intercept ($\mu_3 = -6.33$), indicating that on days when tournament games are not being played, users are unlikely to access soccer content on their mobile device.

We also find, unsurprisingly, substantial variation around these population means, with standard errors for the population distribution in the range of 2-2.5, indicating that some users are quite a bit more (or less) likely to access certain platforms,

($diag(\Sigma_\mu) = (6.358, 4.233, 5.012)$). By contrast, the variances in Σ_e are quite small ($diag(\Sigma_e = (.089, .105, .105, .067)$), indicating that after the user's general propensity to use a platform is accounted for, there is little residual error in daily usage (other than that driven by the aggregate tournament effects).

2.5.2. Correlation Structure

The correlation structure among the channels across time, and within each day is one of the central areas of interest to media planners. As previously described,we summarize such long-run and daily usage effects in two ways. One is the covariance among users' propensities to use each of the platforms over the course of the tournament, which is captured by Σ_μ, and the other is the covariance among an individual's usage of the platforms on a given day, which is captured by Σ_e. The estimates of the two covariance matrices are summarized in Table 4. We observe a strong correlation between ESPN.com and ESPN3 at the long-term level (cross-day posterior mean correlation .795). Thus, heavy users of ESPN.com also tend to be heavy users of ESPN3, which isn't too surprising given that both platforms are accessed with the same type of device, and so all users who access ESPN.com have the opportunity to access ESPN3. (Note that this correlation estimate is only for active users

Table 4: Estimated Model Parameters: Intercepts and Error Structure

Parameters		Mean	2.5% -ile	97.5% -ile
Proportion of Active Users (p_{active})				
p_{active}		**.498**	.455	.536
Population Mean for Platform Intercepts (μ)				
	.com	**-5.01**	-5.01	-4.66
	ESPN3	**-3.81**	-3.82	-3.49
	Mobile	**-6.33**	-6.31	-5.96
	TV	**-2.06**	-2.06	-1.66
Correlation/Variances for Platform Intercepts (Σ_μ)				
Correlation	.com/ESPN3	**.795**	.749	.835
	.com/Mobile	.056	-.065	.171
	ESPN3/Mobile	.028	-.089	.144
Variance	.com	**6.358**	5.279	7.550
	ESPN3	**4.233**	3.326	5.023
	Mobile	**5.012**	3.994	6.340
Correlation/Variances for Daily Error Terms (Σ_e)				
Correlation	.com/ESPN3	-.030	-.385	.329
	.com/Mobile	-.153	-.244	.512
	.com/TV	.083	-.321	.466
	ESPN3/Mobile	.114	-.275	.476
	ESPN3/TV	-.140	-.508	.269
	Mobile/TV	.016	-.404	.436
Variance	.com	**.089**	.051	.149
	ESPN3	**.105**	.062	.172
	Mobile	**.105**	.053	.191
	TV	**.067**	.037	.117

The posterior mean has been highlighted in bold when the posterior interval does not contain zero.

as non-users are "absorbed" by the spike-at-zero; the estimate of the correlation would be higher were the inactive users included.) In fact, knowing that there are some users who use ESPN.com, but not ESPN3, suggests a relatively easy opportunity for ESPN to expand viewership. Interestingly, we do not find a correlation between ESPN.com and ESPN3 at the daily level (corr=-.030). Using ESPN.com *on a given day* does not seem to increase the chance that the user will watch a streaming video on ESPN3.

Interestingly, the relationship with ESPN's mobile platform is quite different, and of great business importance given the recent investments made by ESPN (and many other media companies) in their *mobile* platforms. The mobile channel does not show any significant long-term or daily correlations with any other platforms. This suggests that mobile usage is not cannibalizing usage of the other platforms.

Finally, we are able to estimate the within-day correlations between TV and the other three platforms, even though we do not directly observe which users are watching TV. Furthermore, the posterior intervals for all the correlations between TV and the other platforms contain zero, suggesting that TV usage on a given day is neither positively or negatively correlated with the use of the digital channels. Interestingly, the posterior mean correlation between ESPN3 and TV is -.140 (2.5%-ile = -.508, 97.5%-ile=2.69), suggesting (directionally) that ESPN3 and TV do compete weakly with each other. This is consistent with the fact that TV and ESPN3 offer very similar content (video of full games). By monitoring this parameter over time, as more data is accumulated, ESPN can keep better track of the relationship between ESPN3 and TV, an issue of key business importance.

Summarizing, we find no significant negative correlations between these four channels, suggesting that that the content distribution platforms are not at saturation and that new platforms represent an opportunity to generate incremental reach. This is consistent with ESPN's belief that new platforms do not compete with the old, but allow users to consume media at times that they previously could not. Our finding that the mobile platform seems to provide incremental reach, but is still not the most popular platform is consistent with ESPN's philosophy that users will choose "the best screen available at a given time" (Danaher et al. (2009)).

With our key findings about the relationship between platforms summarized, we now turn to the results that are specific to our case study: the covariates that account for the tournament content.

2.5.3. Tournament effects

As described earlier, the tournament effects include a dummy for whether a given day was on a weekend, the number of games that were played, the number of teams that must "win or go home" on a given day, and dummies for the US team playing, one of three "culturally significant" teams playing (England, Australia or Mexico) and one of the top three teams playing (Spain, Brazil or Netherlands). The posterior summaries for these coefficients are given in Table 5.

Table 5: Estimated Model Parameters: Tournament Covariates

Parameters		Mean	2.5% -ile	97.5% -ile
Weekend	.com	**-.475**	-.712	-.250
	ESPN3	**-.396**	-.628	-.163
	Mobile	.035	-.294	.328
	TV	**.852**	.6385	1.082
Number of Games	.com	**1.370**	.842	1.980
	ESPN3	-.231	-.842	.472
	Mobile	.317	-.388	1.012
	TV	.286	-.190	.705
Number of Teams That Must "Win or Go Home"	.com	.174	-.304	.707
	ESPN3	.095	-.495	.667
	Mobile	-.145	-.817	.566
	TV	.081	-.270	.492
US Team Playing	.com	.338	-.098	.821
	ESPN3	.328	-.081	.741
	Mobile	**.699**	.213	1.186
	TV	**.523**	.227	.935
Canada, Australia or Mexico Playing	.com	.350	-.054	.726
	ESPN3	.028	-.502	.444
	Mobile	.075	-.425	.583
	TV	.037	-.309	.415
Top Team Playing	.com	.211	-.118	.557
	ESPN3	.134	-.314	.563
	Mobile	.073	-.375	.549
	TV	.207	-.116	.544

The posterior mean has been highlighted in bold when the posterior interval does not contain zero.

Our results are consistent with the common notion that people are more likely to watch TV on weekends ($\beta_{14} = .852$) and less likely to go online on weekends ($\beta_{11} = -.475$ and $\beta_{12} = -.396$). (These parameters correspond to users being 2.3 times as likely to watch TV on the weekend and about .6 times as likely to go online on the weekend.) However, we find no weekend effect for the mobile platform ($\beta_{13} = .035$). This provides important insight for media planners; it appears that unlike the other media platforms, mobile is equally accessible and used on both weekends and weekdays. While we can only speculate on how mobile will be used in the future, this lack of a day-of-week effect suggests that media plans for the mobile platform will be different than those for TV and online.

Turning to the tournament content itself, we see a (sensible) significant effect for the number of games played on a given day. When there are a large number of games, interested soccer fans *increase* their usage of ESPN.com substantially as it becomes difficult to follow all the games live on TV, ESPN3 or mobile. Hence, ESPN.com becomes a more attractive platform, while the other platforms are relatively unaffected.

We do not find effects on any of the platforms for the number of teams that must "win or go home." That is, we do not see any evidence that "clincher" games attract more viewership on any of the platforms.[8]

Turning to our set of dummy variables for which teams are playing, we find that all platforms, particularly the mobile and TV platforms are more popular when the US team is playing. The estimated parameters indicate users are 2.0 times as likely to access mobile content when the US team is playing and 1.7 times as likely to watch TV.

Interestingly, and perhaps surprising to non-US soccer fans, we find very weak (but positive) effects when a top team (Spain, Brazil or Netherlands) is playing, suggesting that the American audience we observe is more interested in the US team than these top rated soccer teams. For the variable that measures the aforementioned culturally significant teams, we find a weak positive effect for ESPN.com, but not the other platforms. Thus, we find that this US-based audience seems to be most likely to consume content when the US team is playing and is relatively indifferent to which other teams are playing.

Finally, we should note that we are able to achieve good fit with a relatively simple set of covariates describing how many games are being played and who is playing. Note that there are no covariates that describe the "arc" of the tournament; no dummies for the group

[8]We thank an anonymous reviewer for the suggestion to include this variable and report the null finding because readers may find it interesting.

stage versus the knockout stage, the final game, etc. While we remain a long way from a complete theory of what makes a game attractive to watch on a particular platform, we note that we are able to capture the aggregate viewership (see Figure 2) with a relatively parsimonious set of covariates.

2.6. Forecasting Alternative Media Plans

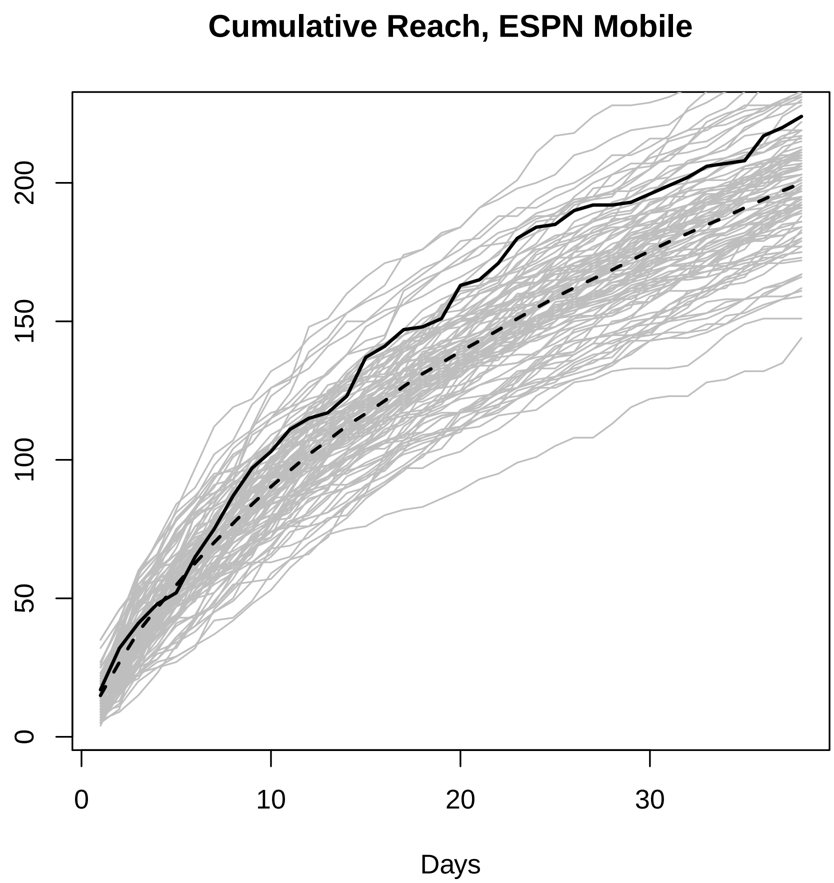
In this section, we present forecasts (counterfactuals) for three alternative media plans[9] that ESPN could have used instead of the large-scale "ESPN XP" program, which provided coverage for every game on all four platforms. First, relating to the key business question of whether it is valuable to invest in coverage on the mobile platform, we created a scenario in which mobile coverage for the tournament was withdrawn entirely, leaving the coverage on the other platform as it was. Figure 6, shows the predicted cumulative reach for the mobile channel (dotted line), which is somewhat lower than the actual reach when there was full mobile coverage (solid line). However, when we look at reach across all the channels, we find no predicted drop in cumulative reach for the tournament (49.7% forecast versus 49.7% actual). This suggests that, at least in 2010, providing mobile coverage did not have a substantial impact on the total number of people that watched the tournament. So, despite the fact that the mobile platform does not seem to be cannibalizing the other platform (as evidenced by the estimates of the correlations between platforms), it doesn't seem to be providing a great deal of incremental reach either, suggesting an interesting counterbalancing between Σ_μ and Σ_e.

Our second counterfactual represents a "compromise" scenario where ESPN provides mobile coverage only on days when the US team is playing. Figure 7 plots the predicted cumulative reach for mobile, had there only been coverage on those four days that the US team played. We find that predictive cumulative reach is not substantially impacted when mobile coverage is reduced; comparing the prediction (dotted line) to the actual cumulative reach (solid line),

[9]We note that these plans are illustrative only and do not represent media plans that ESPN has or might be considering.

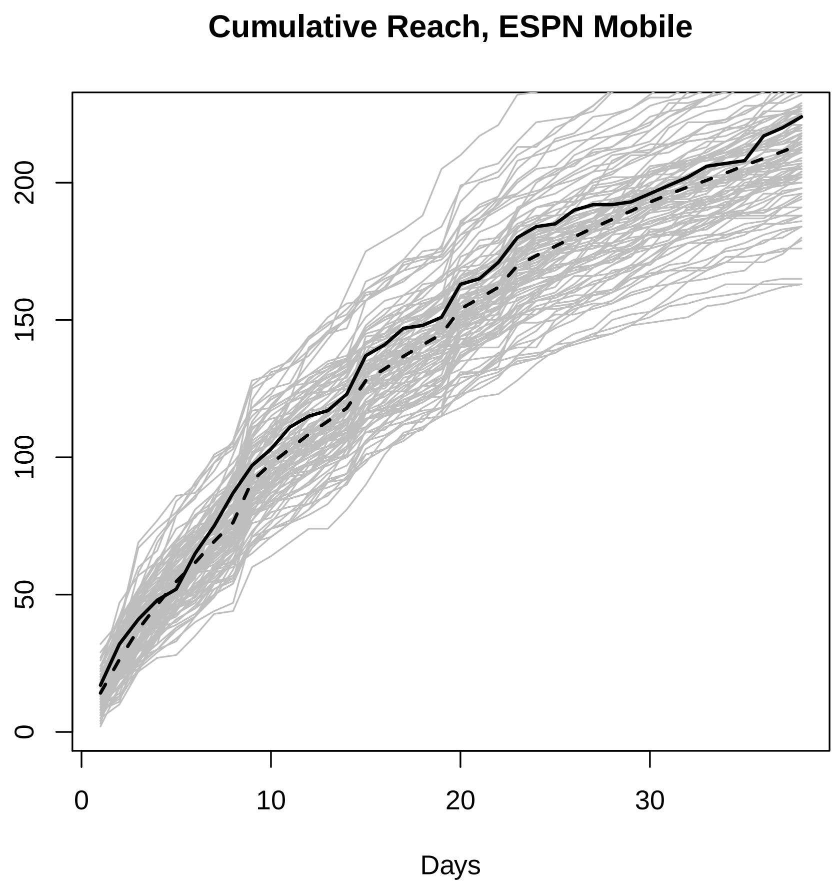
we see that the predicted reach is only slightly lower when the mobile coverage is reduced (and well within the band of prediction error). In total, we predict that 11.0% of the 2,000 users would have watched the mobile platform at all during the tournament compared to the 11.9% we observed in the actual data where there was mobile coverage for all games in the tournament. (By contrast, when we forecast what would have happened had TV coverage been reduced just to the days when the US was playing, we find that reach for television is predicted to be substantially reduced.) This suggests that this lower-cost compromise plan could have allowed ESPN to achieve the same reach for the mobile platform at a substantially lower cost.

Figure 6: Cumulative Reach for the Mobile Platform Had There Been No Mobile Coverage



We draw the forecast for cumulative reach and compare it to the actual statistic (computed directly from the sample) drawn with a solid line. To show the forecast uncertainty, we also draw the prediction for each of 100 random draws from the posterior with grey lines.

Figure 7: Cumulative Reach for Mobile Had There Been Mobile Coverage Only for US Games



We draw the forecast for cumulative reach and compare it to the actual statistic (computed directly from the sample) drawn with a solid line. To show the forecast uncertainty, we also draw the prediction for each of 100 random draws from the posterior with grey lines.

The forecasts presented here represent a small fraction of the types of forecasts that can be done with such a model. Other forecasts that might be of interest include forecasting viewership for different outcomes of the tournament, e.g., "what if the US team made it to the finals?"; albeit this would require integration of the distribution of the set of potential tournament outcomes. If ESPN also had some influence over the tournament structure (which may be more likely for US-based tournaments rather than the World Cup), the model could be used to predict media consumption impact due to changes in the tournament schedule, e.g., "what if the US game was held on a weekend instead of a weekday?" All of these counterfactuals allow us to assess the impact of media planning decisions on the key economically meaningful outcomes: multi-platform reach and exposure.

2.7. Discussion

As the number of media platforms proliferates, complicating the planning problems for media companies, it is important to also appreciate the opportunities that the newer digital platforms offer. The rich, granular data sets that emerge from these platforms allow media companies an unprecedented opportunity to track and model the behavior of individual users over time. The resulting data can be used to help us know ourselves better as consumers and to improve business practice, by modeling the reach and depth of media consumption and the interplay between platforms, as described here. As we have shown, the model we proposed can be used to assess cannibalization between platforms and to forecast media reach for alternative multi-platform media plans. The data fusion approach we describe presents a "way forward" for analysts who have long been thwarted by differing levels of aggregation across platforms. We expect this modeling framework to be applied more generally to other data structures that describe individuals' use of multiple-platforms over time.

With that said, we clearly recognize that this chapter is just a first step towards a comprehensive audience-based measurement and optimization tool. For all that this model encompasses, there are other aspects of media planning not addressed here that would be of both statistical and business interest. We briefly lay out some of these issues for future researchers.

First, while we don't pursue this avenue of research here, we encourage research into what makes games interesting to watch. If we had full understanding of what makes games appealing, we could create a parsimonious model of game attractiveness and use that model to predict viewership. One could imagine combining this model with a game outcome simulator that would allow media planners and buyers the ability to integrate reach and frequency estimates over the predictive distribution of game outcomes, thereby allowing media planners to understand how specific wins and losses might affect media consumption. The ultimate dream for media planners might be a tool that would allow them to answer "what-if" questions such as, "what if the US team makes it to the semi-finals?" Combining our media consumption model with the wealth of game outcome predictions that can be captured from online betting sites, such a simulator might be much more future reality than future dream.

Second, while we know in our case certain characteristics of the soccer content we have investigated (e.g., day-of-week, which teams were playing) seem to influence media consumption overall, a comprehensive forecasting system would take into account a number of additional context effects and interaction effects. For instance, in the World Cup setting, do certain types of teams create a larger audience on certain platforms or with certain types of media consumers? Linked to this issue, do advertisements with certain characteristics work better on high-definition broadcast platforms (e.g., television and ESPN3) than others? We did not explore these issues as fully as we might have, simply because our data was not rich enough to support this type of inference. For example, we explored many other possible specifications for how team attractiveness is related to media usage including a model which estimated attractiveness parameters for each team, but found that the data was insufficient to support estimation by this specification, i.e., the posterior intervals were extremely diffuse. We also considered a specification that allowed for an additional term when two attractive teams played each other, but again found that the data was unable to support estimation of this specification. The World Cup tournament structure simply does not offer enough variation in which teams play on which days to identify these nuances in how the teams that are playing affect media consumption.

There are a number of ways that the data could be enriched to answer some of these important questions, such as observing users over a longer time period on a finer time scale, bringing in other data sources about users or teams as priors or covariates, or adding in data on marketing activities designed to encourages viewership. These richer data sets will be available soon and we encourage others to take up these questions, building on and enriching the modeling framework we have laid out.

We identified some important effects of tournament content on usage that could be important in planning media content. For instance, we found that mobile usage is not lower on weekends (unlike other digital platforms). However, this study stops short of optimally designing tournaments to maximize some outcome that could be predicted by our model (be it reach, GRPs, viewership for the final game, etc.). Of course, for a tournament as internationally important as the FIFA World Cup, it may be a political impossibility to change or influence the tournament schedule, even for a large media player like ESPN; but, there remain many open questions for ESPN such as which games should be shown live, should they advocate to have certain games (such as those involving the US) played on certain days or spaced out over time, etc. Thus, there is an opportunity to combine statistical methods as described here with optimization methods from the Operations Research tradition to identify the best tournament designs to maximize the number of people who watch the tournament. This is related in general to the problem of optimal media scheduling that has been addressed in a number of works (c.f., Danaher and Mawhinney (2001)) but is taken as given here.

We conclude with some learnings from this practical case study and in particular working with potentially rich but messy digital media data sets. First, marketers should become much more connected with the information systems/computer science community as handling large data bases and constructing easily accessible data sets is not in many of our skill sets, but will need to be going forward. Second, there are many “wish list” items one can hope for, e.g., linking to advertisement data, linking to click-throughs on ads, real-time data, etc. While these might be nice ideas conceptually, there are practical limitations to what data we can actually get. But rather than “retreating”, our experience suggests getting the best data you can reliably count on, and in the rich tradition of applied research, model the data“as it lies,” not as you might dream. Data capture has far exceeded data integration and access; however, we believe models like the ones described here should pre-date the existence of large digital media datasets because we know they will be coming soon.

CHAPTER 3 : Information Reweighted Priors

3.1. Motivation and Problem

In current marketing research practice, information arrives at a very rapid pace so that there are no more truly "static" data sets for inference and decision-making. Because technology-enabled marketing research (and information collection) has led to a tremendous speed-up in the flow of information, methods in marketing are needed that allow for coherent sequential information integration in a rapid manner. In practice, relevant external information may "flow in time" from previous research (Lybbert et al. 2007, Higgins and Whitehead 1996), experts (Sandor and Wedel 2001), theories (Montgomery and Rossi 1999), or newly arriving external datasets or resources (Lind and Kivisto-Rahnasto 2008, Lenk and Rao 1990, Putler et al. 1996, Wedel and Pieters 2000, Hofstede et al. 2002), all of which a researcher might want to include in their analyses. But how?

Bayesian inference, which provides a unified approach to modeling with the incorporation of prior information, has become an ever-growing paradigm for statistical inference especially given today's increased computational power. However, for many Bayesian applications, available prior knowledge may be difficult to incorporate into the analyses. Indeed, Bayesian modeling commonly utilizes non-informative or weakly informative priors (Gelman et al. 2008, Gelman 2006) as if external information was not available. In Montgomery and Rossi (1999), prior information on price elasticities is imposed by constructing additive utility models with suitable restrictions on specific parameters. However, such methods are usually application-specific, and not generalizable to a unified system. In other marketing situations, available information may not be readily translatable into informative priors. This might arise when external information is simultaneously related to many parameters, as would occur with information about the relative ranks of future observations, leading to difficulties in coherent prior specification for all the related parameters. It might also arise when external information is on a different scale than the current dataset, such as external information that exists at an aggregated level that corresponds to panel data at the individual level. For instance, the Bayesian analysis of individual-level purchasing/browsing or web traffic data from a web company's service log might be enhanced by aggregate macro records available from industry reports. This is the classic 'data fusion' problem that is pervasive in industry today (Musalem et al. 2008, Musalem et al. 2009). Incorporating such types of information into a marketer's decision making process, where the information is (not easily) translated to a prior on parameters is the kind of problem we would like to address.

We describe in this chapter a new approach to information integration (meta-analysis, Sutton and Abrams (2001), Trikalinos et al. (2008), and an application of meta-analysis with informative priors as in Higgins and Whitehead (1996)) in a model-based setting, an approach we call information reweighted priors (IRPs). In particular, we adapt existing Bayesian methods that have become popular in marketing (Rossi and Allenby 1993) due to their ability to handle heterogeneity, prior information (that we will return to), and allow for shrinkage, by utilizing methods initially developed for fast and efficient computation of case influence deletion (Bradlow and Zaslavsky 1997) and outlier detection (MacEachern and Peruggia 2000) when a Bayesian model has been fit using Markov Chain Monte Carlo Methods (MCMC, Robert and Casella 2004, Gelfand and Smith 1990). The IRP is defined to be a set of informative priors consistent with external information. There are many situations where informative priors might be applied, for example, straightforward prior information about regression parameters or variance and/or covariance parameter (Lenk and Orme 2009), constraints on the parameters (Boatwright et al. 1999), or future observations as in this chapter. Specifically, the IRP approach is a sample reweighting approach, that is characterized by the following "pseudo-code":

(i) Fit an appropriate Bayesian model using current prior knowledge, obtaining a sample from the posterior distribution.

(ii) New information source arrives.

(iii) Reweight the posterior distribution using the IRP approach.

(iv) Step (iii)'s posterior is now step (i)'s output, and when new information arrives return to step (ii).

This pseudo code, while concise, contains many points that are relevant to marketing scholars and practice. First, the overall flow of the pseudo-code suggests the need for inference methods that don't require multiple repeated runs of MCMC methods. This can benefit researchers due to time limitations that are caused by having to run numerous MCMC chains. In contrast, using our IRP approach the MCMC is run once, and then its posterior samples becomes a database (of sorts) for future researchers that can be shared, similar in spirit to multiple imputation methods (Little and Rubin 1983), and sequentially updated.

Second, part of this research discusses step (ii) and (iii), new information sources and how to incorporate them, and this is where the strength and flexibility of our IRP approach lies. That is, standard Bayesian analyses allow for prior/new information integration (by definition), but what has not been addressed, and is novel to this research (as mentioned above), is how to integrate new information when that information is on a different scale than the original model. For instance, imagine a standard hierarchical logit model with heterogeneous slope coefficients. If the prior information is on the slope coefficients, then there is no adaptation needed and the informative prior is put on the slope. But, what if the information about the problem is about some predicted future observation, the rank ordering of a set of betas, etc.? Standard methods are not designed to incorporate that information. The IRP approach does exactly that, in a fast and practical manner, and one that does not require the end user to run MCMC methods again.

To further motivate the potential of the IRP approach, imagine a firm that uses a panel dataset of customer purchases to improve sales and/or promotion strategy for several brands of merchandise by building a predictive model of future outcomes. Suppose external information about future outcomes is available, but in a "non-traditionally used by researchers but ubiquitous form" such as a rank ordering of customers with respect to their future purchase behavior. For example, the marketing researcher knows that customer i is likely to buy more than customer j. In this case, the information is about the ordering of predictions rather than parameters, which leads to difficulty in traditionally incorporating the information by simply choosing a set of prior distributions for person i or person j's parameters as is normally done. As a second example, imagine the external information is on a different scale than the business problem for which inference is desired. For example, the goal of the firm may be to identify the top 30% of the customers (classification) with respect to future purchases, but prior external information about the top 30% is not directly available. Lastly, imagine that the firm's manager knows the probability that each given customer will be the top buying customer in the future, a potentially intuitive and obtainable quantity. We will demonstrate the use of this type of information (the probability of each customer being the top) as a device to construct managerially relevant priors, and hence one way to obtain information to utilize the IRP approach. This is related to an established statistics literature on Urn modeling (Guiver and Snelson 2009), and is described in detail. As will be shown, the IRP approach can incorporate external information of each of these types (and more), demonstrating the generality of the method.

Lastly, step (iv) of our approach is the key meta-analytic computational contribution. We demonstrate through both a real data example and simulation that based on an initial MCMC sample, one can use much lower cost and simpler reweighting methods to incorporate newly arriving information.

The remainder of this chapter is as follows. In Section 2, we describe the IRP approach, which as mentioned has as its core an initial run of an MCMC sampler. Section 3 contains an application of the IRP approach to a central problem in marketing today, advertising effectiveness. We utilize a dataset obtained from Organic Inc, one of the world's largest advertising agencies, combine its information to previously published studies in marketing using the IRP method as a meta-analytic engine, and demonstrate that the information can be integrated sequentially. In Section 4, we demonstrate general properties of the IRP approach using simulation methods where we show how IRP-based inferences improve as new information sources arrive, but potentially at a cost in global model fit. We conclude with thoughts for future research in marketing meta-analytic methods.

3.2. Information Reweighted Prior Method

3.2.1. Definitions and Forms

We start by laying out general notation for hierarchical models that are common in marketing, followed by the basics of the IRP approach. Suppose observations Y (e.g. sales) are assumed to follow a parametric model $p(y|\theta)$, and an 'initial prior' $\pi(\theta)$ on the unknown θ is under consideration. Let Y^* represent unobserved future values, and suppose external information about $G = G(Y^*, \theta)$, a function of parameters and/or predictions, is available. Note that under the initial prior π, $p(y^*, \theta) = p(y^*|\theta)\pi(\theta)$ induces a joint distribution on (G, Y^*, θ) from which distributions of interest such as the marginal $p(G)$, and the conditional distributions $\pi(\theta|G)$ and $\pi(G|\theta)$ can be obtained.

Now suppose the external information about G can be summarized by a distribution $p_e(G)$. To incorporate this information into inferences about θ, we propose the Information Reweighted Prior (IRP)

$$\pi_{IRP}(\theta) = \int_G \pi(\theta|G) p_e(G) dG, \tag{3.1}$$

an update of the initial prior $\pi(\theta)$ with the information carried by the external source $p_e(G)$. Noting that the initial prior can then be expressed as $\pi(\theta) = \int \pi(\theta|G)p(G)dG$, $\pi_{IRP}(\theta)$ in (3.1) is obtained by replacing the marginal $p(G)$ with $p_e(G)$ in this expression.

A useful re-expression of (3.1), by

$$\pi_{IRP}(\theta) = \pi(\theta)\, w(\theta) \tag{3.2}$$

where

$$w(\theta) = \int_G p(G|\theta) \frac{p_e(G)}{p(G)} dG, \tag{3.3}$$

reveals that the IRP update is equivalent to a reweighting of $\pi(\theta)$ by $w(\theta)$, the integrated update of $p(G|\theta)$ in (3.3).

Conditioning on the observed data Y, the posterior $p_{IRP}(\theta|Y)$ update of $\pi_{IRP}(\theta)$ can be expressed as

$$\begin{aligned} p_{IRP}(\theta|Y) &\propto p(Y|\theta)\,\pi_{IRP}(\theta) \\ &\propto p(\theta|Y)\,w(\theta). \end{aligned} \tag{3.4}$$

Analogous to (3.2), (3.4) reveals that the $p_{IRP}(\theta|Y)$ update is equivalent to a $w(\theta)$ reweighting of the initial Bayesian posterior $p(\theta|Y)$. That is, if an MCMC has been run (once) under $\pi(\theta)$, one simply reweights that sample using $w(\theta)$.

It may be of interest to note that $\pi_{IRP}(\theta)$ remains identical to $\pi(\theta)$ when $p_e(G) \propto p(G)$, in which case $p_e(G)$ does not provide more information than what is induced by the original prior. Thus, the IRP approach subsumes standard non-informative situations. For example, this would occur when $\pi(\theta)$ is label invariant and $p_e(G)$ is constant.

Various strategies for adjusting the posterior distribution for additional apriori information have appeared in the extant literature. Arjas and Gasbarra (1996) and O'Donnell and Coelli (2005) utilize posterior selection to adjust for the restrictions without uncertainty on parameters. Bradlow and Zaslavsky (1997) and Peruggia (1997) reweight posterior distributions with the ratio of the new to old posteriors in case influence analyses. Ibrahim and Chen (1998), Ibrahim and Chen (2000) and Chen and Ibrahim (2006) propose a power prior that reweights the original prior with the likelihood of historical data.

In contrast to these strategies, the IRP approach uses an additional probability distribution to capture supplemental external information, and then properly updates the original prior and posterior by reweighting. It is a fully coherent probabilistic approach for incorporating additional external information that is consistent with the Bayesian paradigm, an approach which has not yet been developed in the literature. Also, as we demonstrate, it allows for sequential updating of information as it arrives.

3.2.2. Sampling from an IRP Posterior Distribution

The form in (3.4) suggests that in problems where simulation sampling from the initial posterior $p(\theta|Y)$ is available, importance sampling adjustments based on $w(\theta)$ can be used to compute $p_{IRP}(\theta|Y)$, Owen and Zhou (2000). Such simulation sampling can often be accomplished with MCMC algorithms, Robert and Casella (2004). Assuming for the moment that the value of $w(\theta)$ in (3.3) can be computed for any θ, the following importance sampling methods may be useful.

Suppose $\{\theta_1, \ldots, \theta_K\}$ is a simulated sequence of θ values that is converging in distribution to $p(\theta|Y)$, and let $\{w_1, \ldots, w_K\}$, where $w_k \equiv w(\theta_k)$, be the associated importance weights. Then, to compute posterior quantities of interest such as the posterior expectation of a function $H(\theta)$,

$$E_{IRP}H(\theta) \equiv \int_\theta H(\theta) p_{IRP}(\theta|Y) d\theta, \tag{3.5}$$

the weighted sum

$$\frac{\Sigma_k H(\theta_k) w_k}{\Sigma_k w_k} \tag{3.6}$$

will be a consistent estimator of $E_{IRP}H(\theta)$.

Going further, one can use the idea of sample importance resampling (SIR) to obtain a sample from $p_{IRP}(\theta|Y)$ (Rubin et al. 1988, Smith and Gelfand 1992). Such a sample $\{\theta_1^*, \ldots, \theta_J^*\}$, can be obtained by sampling from $\{\theta_1, \ldots, \theta_K\}$ with replacement according to probabilities proportional to $\{w_1, \ldots, w_K\}$. An attractive feature of such a resample is that it can be used to obtain a sample of predictions $\{Y_1^*, \ldots, Y_J^*\}$, where $Y_j^* \sim p(y|\theta_j^*)$. Note that $\{Y_1^*, \ldots, Y_J^*\}$ is effectively a sample from the IRP predictive distribution

$$p_{IRP}(Y^*|Y) = \int_\theta p(Y^*|\theta) p_{IRP}(\theta|Y) d\theta. \tag{3.7}$$

The simulation sampling above facilitates posterior inference about any function $T = T(Y^*, \theta)$ of Y^* and/or θ that captures an aspect of interest. Once $\{\theta_1^*, \ldots, \theta_J^*\}$ and $\{Y_1^*, \ldots, Y_J^*\}$ have been obtained, simple substitution yields $\{T(Y_1^*, \theta_1^*), \ldots, T(Y_J^*, \theta_J^*)\}$, a posterior-predictive sample of $T = T(Y^*, \theta)$ which can be used for further inference.

Lastly, we address the issue of computing $w(\theta)$ for any given θ, which is necessary for the implementation of the procedures above. For this purpose, consider two cases. When $G = G(Y^*, \theta)$ is unrelated to Y^* so that $G \equiv G(\theta)$, (3.3) reduces to $w(\theta) = \frac{p_e(G)}{p(G)}$ which can be directly computed by substitution or direct evaluation. But more generally, when $G = G(Y^*, \theta)$ depends on Y^*, it will typically be necessary to approximate $w(\theta)$. For example, in settings where it is possible to obtain a simulated sequence $\{G_1, \ldots, G_M\}$ converging in distribution to $p(G|\theta)$,

$$\frac{\Sigma_m [p_e(G_m)/p(G_m)]}{M} \tag{3.8}$$

will be a consistent estimator of $w(\theta)$.

To perform the calculations in the simulations and applications in this chapter, we have used instances of the above methods. Alternatively, numerical approximation methods may prove to be fast and adequate in other problems. Further investigations of computational issues will certainly be of future interest.

3.3. Application: Organic Online Advertising Data

In this section, we apply the IRP method to a real online advertising dataset with external information obtained from previous online advertising studies in one case, and using part of the dataset (not used for calibration) in another. The data were provided by Organic Inc, an online advertising agency, which recorded the banner advertisement exposure, website visit and conversion data for a single automobile brand during a nine week online advertising campaign. The data are at the individual-level, and each observed user is recorded when he/she is exposed to an advertisement for the brand (banner advertising), clicks through a search link (search engine marketing (SEM)), or engages in a conversion activity at the website. A successful conversion was determined as traffic to a specific website page where the user finds a car quote, builds and prices a model and/or finds a dealer. In the remainder of the chapter, we refer to this dataset as the 'Organic dataset'. Online advertising has been widely studied (Dreze and Hussherr (2003), Manchanda et al. (2006), etc.), including Braun and Moe (2011), who studied this particular dataset, but for a different purpose. Previous knowledge can be incorporated into the current study with the IRP method. In this application, we focus on a basic model where the effects of the number of advertisement impressions, click-throughs, sites and pages visited on the successful conversion rates are estimated, and show the effect of external information on parameters and predictions.

To verify the model and illustrate the impact of external information, part of the observations are utilized as a calibration dataset and the rest are held out as test dataset to compare with the predictions from the benchmark model, and measure the IRP approach performance. The following 3 models are fit to the calibration data.

In Section 3.3.1, a basic (benchmark) Bayesian logistic regression is conducted with standard priors. This acts as step (i) from the previously mentioned pseudo-code where the "original" MCMC sampler is obtained. In Section 3.3.5, the external information about the sign of the effects of number of advertisement impressions, click-throughs, sites and pages visited from literature is utilized, and applied to the basic Bayesian logistic regression model through the IRP method. The information sources include both academic and business articles that study online advertising. In Section 3.3.5, the external information about predictions is applied to the basic model through the IRP method. Since external information about predictions is not available externally for this dataset because of its uniqueness, we

”construct” external information that was estimated from part of the training data that was not used for calibration[10].

3.3.1. A Bayesian Logistic Regression Applied to Organic Data

We consider a benchmark hierarchical Bayesian logit choice model , where the indicator of successful or failed activities (y_{it}) of customer i at time t is regressed onto multiple independent variables (X_{it}), with the logit link function and heterogeneous coefficients β_i.

Specifically, the panel data we consider are indicators of successful conversion of n customers, over a sequence of T time periods. Letting y_{it} denote the success or failure of customer i at time t for conversion, $i = 1, \ldots, n$, $t = 1, \ldots, T$, the distribution of these choices under the logit model is given by,

$$\begin{aligned} &y_{it} = 1 \text{ with probability } p_{it}, \\ &logit(p_{it}) = X_{it}\beta_i, \end{aligned} \tag{3.9}$$

where $\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T$ is an individual-level parameter vector.

To keep the model concise, we consider four important factors that affect the probability of a successful activity: (i) number of advertisement impressions, (ii) click-throughs, (iii) sites visited and (iv) pages visited in the 7 days proceeding the date of the j^{th} activity of customer i [11]. These four numbers as well as an intercept become the elements of X_{ij}, and consequently the length of the coefficient vector β_i is $p = 5$.

We further assume that the coefficient parameters β_i follows a hyper-structure.

$$\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T \sim \text{Normal}(\bar{\beta}, V_\beta), \tag{3.10}$$

[10] In practice, we do not recommend using part of the dataset to generate external information, since it decreases the number of observations for model fitting. We use the external information generated from part of the dataset just to provide reasonable prediction information and demonstrate the IRP method.

[11] The results are robust to the exact number of days, but discussions with Organic Inc determined the time window.

with mean vector $\bar{\beta}$ and covariance matrix V_β.

As the benchmark, the Bayesian analysis is conducted with the 'standard' diffuse prior distributions as in Rossi et al. (2005), corresponding to the simulation in Section 3.4.

$V_\beta \sim Inverse - Wishart(\nu, W)$, where $\nu = 2 + 5 = 7$ and $W = \nu I_p$;

$(\bar{\beta}, vec(\Delta))|V_\beta \sim Normal(0, V_\beta \otimes 100 I_p)$.

The model is fit to the dataset with two MCMC chains of length 30,000 iterations including 10,000 burn-in iterations. For each conversion in the hold-out observations, we then predict whether a conversion is successful or not. The average false prediction rate is computed as the measure of model performance. The benchmark Bayesian model with non-informative hyper-priors yields 32.29% false prediction rate with standard error 0.49%.

3.3.2. IRP with Parameter Information

In this subsection, we focus on the effect of the number of impressions, click-throughs, sites and pages visited on customer conversion rate, and utilize literature from both academic research and business articles for external information, i.e. to construct $p_e(G)$. Since no previous dataset being researched is exactly the same as the Organic dataset (nor would it ever be in practice), the external information focuses on the signs of the effects, but clearly other summaries are possible. The studies we utilized are listed in Table 6 below. The first two articles in Table 6 are academic papers and the rest are business articles. Assuming the results in Table 6, we utilize these previous study results to 'build' the external information to conduct two different "counterfactual IRP analyses". In the first one, we take the place of a hypothetical manager sitting here today, looking back on ALL of these studies, and incorporating their information into the analyses along with the observed current Organic data. The second analysis we ran was a "back-in time counterfactual meta-analysis" where we assume that the Organic data was observed prior to all of the papers in Table 6 and we sequentially update the Organic data posterior inferences (in real-time) as the new studies come in. We ran both of these studies to highlight that the IRP approach can be used for both batched (counterfactual one) and sequential (counterfactual two) meta-analyses.

Table 6: Literature Review

Effect on Conversion Rate	# impressions	# click-throughs	# sites	# pages
Manchanda et al. (2006)	> 0	effect not significant	> 0	> 0
Moe and Fader (2004)	NA	NA	> 0	NA
Song (2001)	> 0	NA	NA	NA
Briggs (2001)	NA	effect not significant	NA	NA

3.3.3. Batched Meta-Analysis of Organic Data

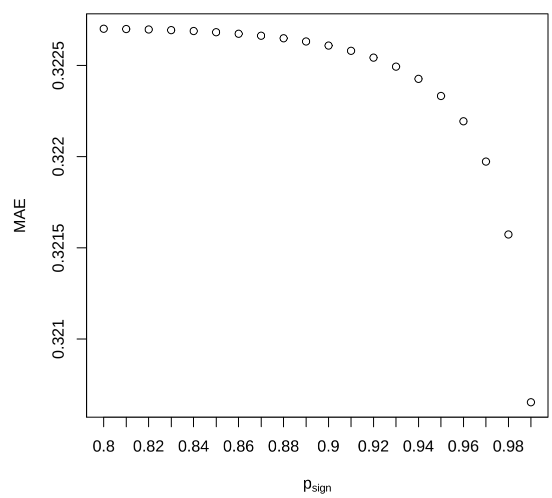
We ran an analysis to fuse together the Organic data set with all of the papers contained in Table 6. In particular, and given that the nature of the external information is limited to signs of coefficients (positive or negative), we ran our IRP approach under a variety of degrees of uncertainty assumed for each study. This is for three major reasons. First, as stated, there are times where external information just comes in terms of positive/negative, but one might not want to assume its sign with certainty. In our case we let $p_{sign,impressions}$, $p_{sign,click-throughs}$, $p_{sign,sites}$, and $p_{sign,pages}$ range from 0.80 to 0.95, a large and realistic range. This is in contrast to extant research that would impose a hard constraint, $p_{sign} = 1$. Second, it is important for sensitivity analyses where the analyst might want to know how the posterior inference of interest changes as the new information is sequentially included (but with error). Note that in cases where the extant research contains a p-value, then that can be used as the degree of uncertainty directly (hence the tight link to meta-analyses). Third, it provides "proof" that the IRP approach can be run easily and quickly under a variety of conditions, thus allowing the use a range of informative priors.

To compute $p_{IRP}(\theta|Y)$ here, where θ represents the set of all parameters in model 3.9, we proceed with importance sampling as described in Section 3.2.2 where $p_e(G) = p_e(\bar{\beta})$ is the prior above induced by previous studies, and $p(G)$ is the prior on $\bar{\beta}$. Then, based on the simulated set of parameters $\{\theta_1, \ldots, \theta_K\}$ drawn from $p(\theta|Y)$ under an uninformative prior, the importance weights $w_k = \int p(G_k|\theta_k)\frac{p_e(G_k)}{p(G_k)}dG$ equals the ratio

$\frac{p_{sign,impressions}^{\bar{\beta}_{impressions}>0}\, p_{sign,click-throughs}^{\bar{\beta}_{click-throughs}>0}\, p_{sign,sites}^{\bar{\beta}_{sites}>0}\, p_{sign,pages}^{\bar{\beta}_{pages}>0}}{0.5^4}$, where 0.5^4 in the denominator comes from a uniform 50-50 prior on each of the coefficients.

Letting the coefficients of the number of impressions, sites and pages to be positive with the same probability p_{sign}, the false prediction rates, out-of sample (i.e. mean absolute error MAE, (Sheiner and Beal 1981)) are drawn in Figure 8. The improvement using the external information compared to the benchmark Bayesian model is considerable, and it increases with a larger p_{sign} which corresponds to a stronger external information signal, suggesting strong evidence of positive effects. It is straightforward to impose different probability p_{sign} in the batched model, as in Section 3.3.4.

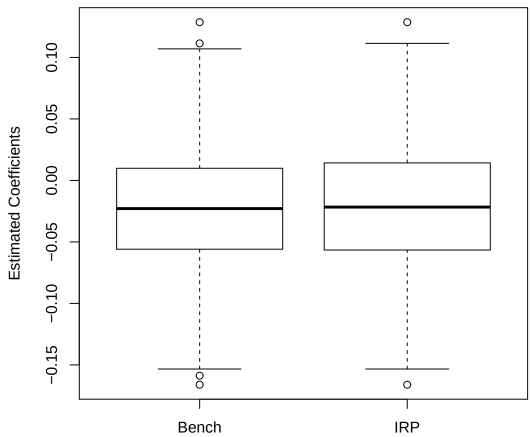
Figure 8: False Prediction Rates for Organic Dataset with Benchmark Bayesian Analysis and IRP



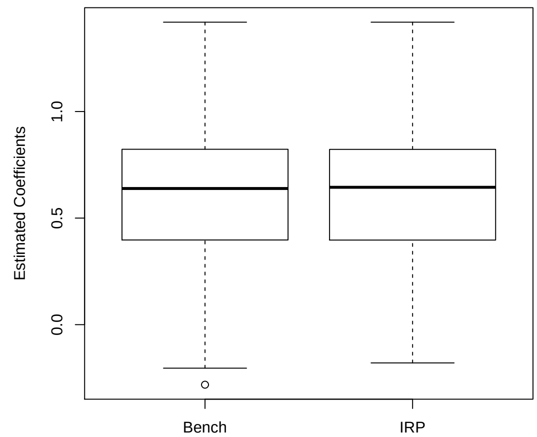
We further explore how the posterior distribution is 'skewed' by the external information. Since the external information favors the case where the coefficients of the number of impressions, sites and pages are positive, it is expected that the posterior distributions are skewed towards the positive domain. As an example, we consider the case where $p_{sign} = 0.9$. The results for the coefficients of impressions and pages are drawn in Figure 9(a) and 9(b).

For the coefficient of number of impressions (Figure 9(a)), the effect of incorporating IRP is notable. Clearly the distribution of the coefficient of number of impressions is skewed upward, and the confidence internal is narrower because the external information shrinks the posterior distribution. The effect of the external information on the coefficient of number of pages (Figure 9(b)) is not as dramatic. From the posterior sampling under the benchmark Bayesian analysis, more than 99% of the samples yield a positive coefficient for the number of pages, which implies the 'signal' of the sign of the effect of number of pages in the data is very strong. Hence the effect of the informative prior is overwhelmed by the signal in data, and the effect of the IRP approach is limited appropriately, and if at all, it shrinks the results (possibly erroneously) 'inward' toward 0 since $p_{\text{sign}} << p_{\text{MCMC}}$. The different effects of IRP on these two parameters show that the effect of external information fades away with stronger signal in the data, which is a desired property in Bayesian analyses. Research on this property of Bayesian analysis can be found in, for example, Lenk and Orme (2009) discuss the effect of priors with sparse dataset, Lenk et al. (1996) consider Bayesian analysis with small samples, and DeSarbo et al. (2010) discuss the effect of priors with ill-defined parameters.

Figure 9: Posterior Samples for Coefficients of Number of Pages and Impressions of Organic Dataset



(a) Coefficient of # Impressions



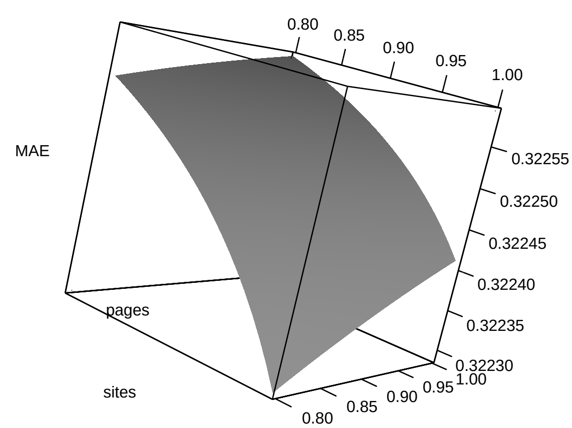
(b) Coefficient of # Pages

3.3.4. Sequential Meta-Analysis of Organic Data

In contrast to Section 3.3.3, we ran a sequential analysis using the IRP approach assuming the appearance of the literature in Table 6 in three stages, where we incorporated each piece of information sequentially as described in the 'psuedo-codes' in Section 3.1. In the first stage the external information implies that the effect of # impressions is positive as in Song (2001); in the second stage, the information implied by Moe and Fader (2004) further suggests the effects of both # impressions and sites are positive; and in the third stage, the information implied by Manchanda et al. (2006) suggests the effects of # impressions, sites and pages are all positive. Setting $p_{sign} = 0.95$, the MAE of predictions is decreased by 0.014%, 0.105% and 0.121% at each of the stages.

Further it is straightforward to implement different levels of uncertainty for each piece of information and look at the bi-variate or tri-variate distribution of information incorporation. For example, letting the coefficients of # of sites and pages vary from 0.8 to 1, one can draw the surface of the MAE as in Figure 10. With IRP, implementing this bi-level external information is quick and does not require extensive computation. The graph is suggesting the information on the number of websites is important to reduce the MAE, while the information on the number of pages leads to a worse MAE. By considering different level of uncertainty for each piece of information, this graph is performing a sensitivity analysis of the effect of external information on the inference.

Figure 10: False Prediction Rates Surface for Organic Dataset



3.3.5. IRP with Prediction Information

To further highlight the power of the IRP approach, we ran a second form of meta-analysis using information about future predictions (Y^*), as motivated in the introduction, because this form of prior information is not simply incorporated into a standard Bayesian analysis. However, unlike the coefficient analysis in Section which we based on extant literature, external information is not available for this dataset about future predictions. Therefore to implement this in a thoughtful way, and construct $p_e(G)$ for Y^*, we split the Organic dataset into three non-equal slices for external information generation, calibration, and out-of-sample validation. To obtain the external information, 1/10 of the training data were randomly chosen to estimate the conversion rate. A exploratory analysis shows that the conversion rate of the chosen data is 0.300 with standard error 0.024. Since this knowledge is imprecise, and based on approximate normality of proportions, we assume that the conversion rate in the test dataset is larger than 0.25 with probability 0.9 (2 standard errors below the mean), and this is utilized as the external information. Thus 1/10 of the data was used for prior construction (other methods are discussed below), 7/10 of the data are utilized for the Bayesian inference calculation and posterior distribution sampling, and the remaining 2/10 of the data are saved for out-of-sample model performance testing.

To compute $p_{IRP}(\theta|Y)$ here, we proceed again with importance sampling as described in Section 3.2.2 where now $p_e(G) = p_e(Y^*)$ is the prior induced by the above information, i.e. $p_e(G) = p_e(Y^*) = 0.9$ if the conversion rate estimated from Y^* is larger than 0.25, and 0.1 otherwise. With this information about predictions, the model yields 29.67% false prediction rate with standard error 0.37%. This is significantly ($p < 0.05$) less than the 32.29% false prediction rate by the benchmark Bayesian analysis.

3.4. A Simulation Experiment

3.4.1. Simulation Setup

To further illustrate the properties of the IRP approach, we describe here a large simulation based on the example introduced in Section 3.1. We consider the analysis of panel data generated by a multinomial logit (MNL) model, a widely used model for customers who ‘pick 1 out of J choices’, which is similar to the logit model specified in Section 3.3.1. The MNL model has been successfully applied, for example, in products sales research (Buckley 1988), product consumption research (Yildiz Tiryaki and Akbay 2010), classification of financial policy (Dubas et al. 2010), brand choice research (Allenby and Rossi 1991) and voting research (Dow and Endersby 2004). See Rossi et al. (2005), Chapter 5 for an excellent overview.

Specifically, the panel data we consider are the purchase choices of n customers, from J

possible brands, over a sequence of T_1^i time periods for the i^{th} customer. We further assume that managerial interest concerns the choices of these customers over T_2 (or T_2^i) future time periods. Letting y_{it} denote the choice of customer i at time t, $i = 1, \ldots, n$, $t = 1, \ldots, T_1^i$, the distribution of these choices under the MNL model is given by,

$$
\begin{aligned}
&y_{it} = j \text{ with probability } p_{itj} \text{ for } j = 1, 2, \ldots, J, \\
&(logit(p_{it1}), \ldots, logit(p_{itJ}))^T = X_{it}\beta_i, \qquad (3.11)
\end{aligned}
$$

where X_{it} is a $J \times p$ design matrix, and $\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T$ is an individual-level parameter vector of length p.

Customer preferences for brand j over brand 1 are captured here by the brand-specific intercepts that comprise the first $J-1$ elements of β_i, namely $\beta_{i,j-1}$, $j = 2, \ldots, J$. Correspondingly, the $j-1^{th}$ column of X_{it} is $[0, \ldots, 0, 1, 0, \ldots, 0]$ with 1 in the j^{th} position. The remaining columns of X_{it} are covariate values that may influence purchases, such as prices, promotions, shelf-space, etc., commonly collected from the marketing domain.

We further suppose that the individual-level coefficients in β_i depend on individual covariates, such as demographic characteristics including household income, family size, etc., as is standard in hierarchical Bayesian models (Rossi et al. 2005). Representing these covariates with the design matrix Z_i, we assume the random effects formulation

$$
\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T \sim Normal(\bar{\beta} + Z_i\Delta, V_\beta), \qquad (3.12)
$$

where Δ is the $p_Z \times p$ coefficient matrix, $\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)^T$ is the coefficient intercept vector, and

$V_\beta = diag(\sigma_1, \ldots, \sigma_p) \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1p} \\ \rho_{12} & 1 & \ldots & \rho_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{1p} & \rho_{2p} & \ldots & 1 \end{pmatrix} diag(\sigma_1, \ldots, \sigma_p)$ is the coefficient covariance matrix.

In this simulation, we focus on identifying future heavy volume buyers of a specific brand, which is of managerial interest to conduct a targeted marketing strategy for loyalty program enrollment (Sharp and Sharp 1997) or for catalog distribution (Shim and Mahoney 1992), to name a few. For a specific brand j, the object of interest here is $T(Y^*, \theta) = S^A_{top}(Y^*)$, the set of customers whose purchase volume across the future T_2 time periods will fall in the top $A\%$ of all customers (i.e. a standard heavy-user classification problem). Just as in the first example, a standard Bayesian analysis will provide the posterior distribution of this set for the purpose of IRP inference that will be reweighted.

3.4.2. Generating the Data

To make the simulations more relevant, we generated datasets from a model which was used to fit a 'classic and general' marketing panel dataset that was studied by Allenby and Rossi (1991). To obtain 'realistic' underlying parameter values for the model, we used the estimated parameter values from the Allenby and Rossi (1991) data.

The dataset considered by Allenby and Rossi (1991) records purchase choices of margarine from 10 brands by 517 households and, as they demonstrate, is fit well by a MNL model. Proceeding as in Chapter 5.4 of Rossi et al. (2005), we based our simulation setup on a subset of that data restricted to $m = 6$ brands, and households restricted to 5 or more purchases. These restrictions reduced the size of the dataset to $n = 313$ customers (households) over 5 to 40 time periods. The customer β_i coefficients for the MNL model (equation (3.11)) consist of the brand-specific intercepts for brands 2 to 6 relative to brand 1, plus a coefficient β_{i6} for the logarithm of price, a covariate of direct interest in our model. Thus, each β_i here is a vector of length 6.

To complete the specification of the data generating model for our simulations, the hyper-parameters in (4.8), namely $\bar{\beta}$, Δ,V_β, were set equal to the estimates from an MNL model fit to this subset of data. With this specification, we simulated the customer β_i's from

(4.8), followed by the customer purchase y_{it}'s from (3.11). We then repeated this simulation to generate 1000 paired sets of β_i's and y_{it}'s. This data was then used to compare the performance of a benchmark Bayesian approach using an uninformative prior, with the performance obtained when additional external information is incorporated (as discussed next) using the IRP approach.

For the benchmark Bayesian approach, we used the relatively diffuse normal inverse-Wishart prior as in Rossi et al. (2005) for the model hyperparameters in (4.8):

$$V_\beta \sim Inverse - Wishart(\nu, W),$$

where $\nu = 6 + 3 = 9$ and $W = \nu I_6$, and

$$(\bar{\beta}, vec(\Delta))|V_\beta \sim Normal(0, V_\beta \otimes 100 I_6).$$

3.4.3. Collecting External Information with an Additional PL Model

To construct an informative prior about the top $A\%$ of the customers, we turn to the statistics literature and a well-established set of models that can generate a distribution of unit ranks. In particular, suppose that external prior information (described below) is available about $(R_1^*, \ldots, R_n^*)$ where R_j^* is the index of the customer with the j^{th} largest future purchase volume across the future T_2 time periods. Suppose also that this prior information can be captured by treating $(R_1^*, \ldots, R_n^*)$ as a sample drawn without replacement from an urn with probabilities $(p_1^e, \ldots, p_n^e)$, the so-called Plackett-Luce (P-L) model (Guiver and Snelson 2009). Thus, for the sequence of draws $(i_1, \ldots, i_n)$,

$$\begin{aligned} & p_e((R_1^*, \ldots, R_n^*) = (i_1, \ldots, i_n)) \\ & = \frac{p_{i_1}^e p_{i_2}^e \cdots p_{i_n}^e}{(1-p_{i_1}^e)(1-p_{i_1}^e - p_{i_2}^e)\cdots(1-\sum_{j=1}^{n-1} p_{i_j}^e)} \end{aligned} \tag{3.13}$$

under this model. Such information could be obtained(in practice) by having managers assign weights to customers which reflect their relative beliefs about which customers will be the highest volume purchasers; but this reflects but one option to obtain them. It might helpful for the manager to note when eliciting values, that p_i^e will be proportional to the probability that the i^{th} customer's future purchase volume will be the largest. It is also possible that these values can be obtained from past data, just noting the fraction of times in past panels in which unit i is the highest, which is commonly done in urn models. It is important to note that the external information here concerns $G(Y^*, \theta) = (R_1^*, \ldots, R_n^*)$ which is not the same as the aspect of interest $T(Y^*, \theta) = S_{top}^A(Y^*)$. That the IRP approach can translate information from one scale (the urn ranking scale if you will) to another (the top A% of the customers) is an attractive feature.

To compute $p_{IRP}(\theta|Y)$ here, we proceed again with importance sampling as described in Section 3.2.2 where now $p_e(G) = p_e(R_1^*, \ldots, R_n^*)$ is the prior above induced by the P-L model, and $p(G)$ is the prior on $(R_1^*, \ldots, R_n^*)$ induced by the standard prior $\pi(\theta)$. Based on the simulated sequence $\{\theta_1, \ldots, \theta_K\}$ from $p(\theta|Y)$ under the standard model, the importance weights $w_k = \int p(G_k|\theta_k)\frac{p_e(G_k)}{p(G_k)}dG$ themselves are each obtained by importance sampling, rather than direct evaluation, because the external information here concerns predictions rather than parameter values. Specifically, for a simulated sequence $\{G_{k_1}, \ldots, G_{k_M}\}$ converging in distribution to $p(G_k|\theta_k)$, w_k will be estimated by $\frac{\Sigma_m[p_e(G_{k_m})/p(G_{k_m})]}{M}$ as in (3.8).

The P-L external information prior given in equation (3.13) on the purchase ranks $(R_1^*, \ldots, R_n^*)$ for this application requires the specification of $(p_1^e, \ldots, p_n^e)$, where p_i^e is the expert's prior probability that the i^{th} customer will be the largest total volume purchaser. Without loss of generality, we focus on brand 2, and its purchases, across T_2 future time periods. For the case of unbiased external information, we here imagine an expert whose knowledge about p_i^e is equivalent to the information provided by $\beta_{i,1,expert} \sim Normal(\beta_{i1}, \sigma_e^2)$, where β_{i1} is the true intercept coefficient for brand 2. Substituting the $\beta_{i,1,expert}$ values for β_{i1} in the true coefficient vectors, we then repeatedly simulated the number of purchases of brand 2 over the next T_2 time periods for every customer. The expert's value of p_i^e is then estimated by the relative frequency with which customer i is ranked first over these repetitions. [12] For the case of biased external information, we imagine an expert who generates independent $p_i^e \sim Unif(0,1)$, a manifestation of 'pure ignorance'. Note that such biased external information is different from vague information. Whereas the latter has large uncertainty, giving flat weight across all posterior samples, the former assigns different weights to posterior samples, weights which may be inconsistent with the truth and skew the priors in an 'undesired direction'. We use such biased information to examine the robustness of the IRP method.

3.4.4. Simulation Results

For each of the 1000 sets of simulated y_{it}'s described in Section 3.4.2, we ran the benchmark Bayesian analysis, and then reweighted the posterior according to the following IRP priors: the unbiased IRP priors with σ_e = 0.05, 0.1, 0.3, 1, 25, 100, and the biased IRP prior. To obtain the posterior samples for the benchmark Bayesian analysis, two MCMC chains were run with 30,000 iterations, and the posterior samples drawn after 10,000 burn-in iterations. The convergence of the MCMC chains was tested with the Gelman and Rubin convergence diagnostic (Gelman and Rubin 1992) to ensure that the samples are drawn after the chains reach convergence.

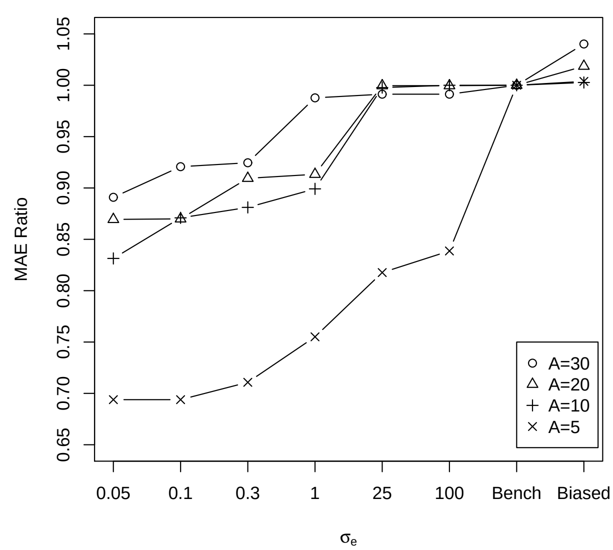
The posterior under each IRP prior induces a set of posterior probabilities $p_{i,A}^{post} \equiv p^{post}(i \in S_{top}^A),\ i = 1, 2, \ldots, n$, which can be computed by simulation. These posterior probabilities can be regarded as estimates of the true set of probabilities $p_{i,A}^{true} \equiv p^{true}(i \in S_{top}^A),\ i = 1, 2, \ldots, n$ for each of set of simulated true β_i's, probabilities which can be computed by repeated simulation of the purchases. To evaluate the effectiveness of the IRP priors for the identification of $S(\theta)$, we assess the overall accuracy of the $p_{i,A}^{post}$s as estimates of the

[12] For the sake of computation ease, if the estimated p_i^e is 0 or 1, it is jittered with a small number (1/10 of the smallest p_i^e or $1 - p_i^e$ where $p_i^e \neq 0$ *or* 1), and hence the posterior samples would not receive 0 weight in the reweighting stage. A sensitivity study about the jittering value was conducted, where the jittering value was ranged from 1/100 of the smallest p_i^e or $1 - p_i^e$ where $p_i^e \neq 0$ *or* 1 to 1/5, and the inference results were not observably affected.

$p_{i,A}^{true}$s by the MAE $\frac{1}{n}\Sigma_{i=1}^{N}|p_{i,A}^{post} - p_{i,A}^{true}|$. For the various priors under consideration and various choices of A, Figure 11 displays the average ratio of this MAE for the IRP posterior estimates to this MAE for the benchmark Bayesian posterior estimates.

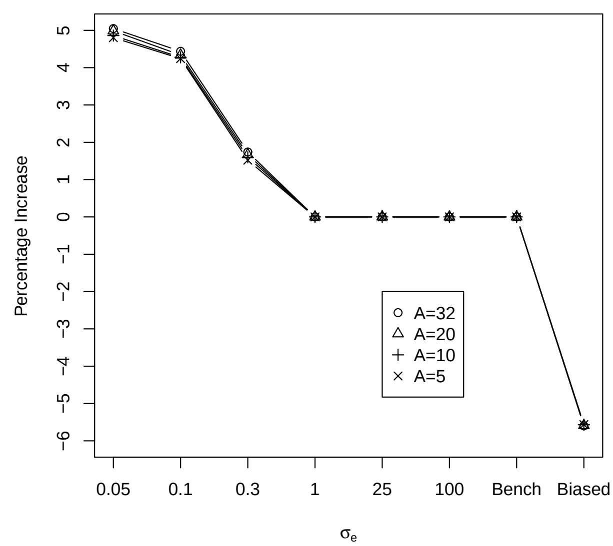
Figure 11 shows that estimation accuracy of the $p_{i,A}^{post}$s is improved by incorporating unbiased information via the IRP. For each value of $A\%$, this improvement increases as σ_e gets smaller and the external information becomes more precise, just as one would expect. More subtly, for each value of σ_e, this improvement increases as $A\%$ gets smaller as the prior probabilities of customers being top ranked (the values of p_i^e) is more relevant for identifying the very highly ranked customers. As $A\%$ gets larger, the actual distribution of the top $A\%$ is further from the distribution implied by the P-L model which is the nature of the external information. Nonetheless, the improvement at all levels of $A\%$ shows that the IRP method can incorporate external information 'loosely related' to the aspect of interest. Use of the IRP with biased information is worse than the benchmark approach which uses using no external information, though only slightly worse.

Figure 11: Ratio of MAE for Top $A\%$ Customer Identification with IRP and the Benchmark Bayesian Approach



We next turn to an evaluation of the IRP method in terms of its potential to increase firm profitability, an issue of direct managerial interest in the marketing domain. We consider what would happen if the firm were to target the future heavy volume purchasers for a campaign designed to increase their sales. More precisely, we suppose the firm runs a marketing campaign targeted at the top $A\%$ customers, a campaign expected to increase the group's purchase by 10%. By simulating from the true underlying distribution of future purchases, we can calculate the total sales increase for any particular targeted group identified by a posterior distribution. To compare the sales increases obtained using the various posteriors for group identification, Figure 12 depicts the percentage difference in revenue from using identifications based on the IRP priors versus the benchmark prior.

Figure 12: Percentage Revenue Increase with Varying σ_e Compared with Benchmark Bayesian Analysis



It is clear from the plot that the largest increase in sales occurs for the smallest σ_e, when the unbiased information is most precise. As σ_e increases, so that target identification gets worse, the sales increases decline stabilizing when σ_e offers no improvement over the benchmark prior. Although the improvement is larger when $A\%$ is large, the improvement is very slight suggesting that the firm will be better off targeting a smaller $A\%$ if the campaign is costly.

3.5. Discussion

The IRP is a unified procedure for meta-analysis subsumed under a Bayesian inference structure. The likelihood, aspect of inference interest and external information are not restricted and hence the method is flexible and generalizable. Our simulations demonstrate that incorporating unbiased information through IRP can improve inference about a related aspect of interest. However, IRP may introduce bias in the inference of other parameters and hence the global fit, as explained in Section 3.2.1. The informative prior approach could involve a trade-off between global fit and a specific business problem, and the choice of the researcher/manager depends on the objective of the specific study.

Despite our research progress, there are several other questions remaining. First, to what extent does the external information affect the estimation of the specific aspect of interest? The IRP utilizes information through the prior distribution, which could be restrictive when there is a large dataset, since the strong signal in data may overwhelm the effect of external information. While this is a desired property of Bayesian analyses, this property restricts the potential effect of the external information. A more generalized method which enables the aspect of interest to play a key role in the model specification is desired for taking into consideration the 'localized inference' more directly, possibly through a tailored likelihood.

Second, how to obtain and process the information source for the meta-analysis is not yet fully addressed in this study. In the application in Section 3.3, we obtain the external information in two ways: (1) from previous literature and (2) using part of the dataset. A certain degree of uncertainty is added to the obtained external information, since the datasets studied in the previous literature are not exactly the same as the dataset under research, and also the information obtained from part of the dataset may not precisely reflect the truth. A more rigorous external information process system is important to fully develop; but for now, computing posterior distributions under a range of uncertainty levels is recommended.

For computation, a simpler and more precise algorithm to sample from the posterior distribution may help improve the efficiency of the algorithm. In this study, we utilize importance sampling, which might be 'expensive' if the computation of the weights is not straightforward. The proposed distribution can certainly be improved for higher efficiency as mentioned in Section 3.2.1, and other sampling techniques may be a subject of future research. We believe this chapter is a good first step, for those looking to incorporate prior information on scales that are readily available, and are likely to exist in practice; for example the ways managers think as opposed to thinking on a parameter scale in which most priors are constructed.

CHAPTER 4 : Rank Enhanced Likelihood

4.1. Motivation and Problem

Bayesian inference, which provides a unified approach to modeling and incorporating prior information, has become an ever-growing paradigm for statistical inference due to increased computational power. In most Bayesian applications, the likelihood is selected to focus on global fit. Yet in some cases, practical problems involve inference at the individual-level, such as ordering, for which overall model fit is not sufficient as a measure of model appropriation.

In such cases, one possible solution is to incorporate prior knowledge from previous or external research (Lybbert et al. 2007, Higgins and Whitehead 1996), experts (Sandor and Wedel 2001), theories (Montgomery and Rossi 1999), or external datasets or resources (Lind and Kivisto-Rahnasto 2008, Lenk and Rao 1990, Putler et al. 1996, Wedel and Pieters 2000, Hofstede et al. 2002). However, incorporating external information could be restrictive when there is a large dataset, since the strong signal in data may overwhelm the effect of that information. This chapter proposes a more generalized method which enables the information (in our case the ranking information) to play a key role in the model specification, where the main idea is to consider a model that puts a likelihood on the ordering as opposed to it being just an outcome of the process for the observed Y's. We call it the 'rank enhanced likelihood' (REL). This method poses no restrictions on the initial likelihood, prior or hyper-prior distributions, or data structure, hence is a very general method for marketing scientists and researchers.

To motivate the REL method, we consider one commercial example as described below, where ranking is the inferential goal. Imagine a firm selling several brands of merchandise that utilizes a panel dataset to improve its sales and/or promotion strategy. In practice, the ranking of the market share of the brands is crucial for branding and marketing in terms of understanding optimal resource allocation of the marketing budget. As opposed to modeling only y_{ijt} as the sales of brand j at time t to customer i and ranking based on μ_j the population mean, we demonstrate that incorporating a ranking part to the likelihood directly may improve the inference for ranking. This example is further studied with simulation in Section 4.3. Finally, to demonstrate the effectiveness of the REL approach in real-world problems, a grocery sales scanner dataset collected by the now-defunct ERIM division of A.C. Nielsen on panels of households in two mid-sized Midwestern cities, is studied.

The rest of the chapter is organized as follows. In Section 2, we define the REL and propose a procedure to sample from the corresponding posterior distribution. In Section 3, we study the examples illustrated in this section, and conduct simulations to compare the performance of REL and standard Bayesian analyses. In Section 4 we apply the REL method to the ERIM dataset. In Section 6 we conclude with a discussion of future research.

4.2. Rank Enhanced Likelihood Method

4.2.1. Motivation and Definition

Suppose a panel dataset of K objects (say, sales data of brands as above described) $Y = \{y_{kt}, k = 1, 2, ..., K, t = 1, 2, ..., T\}$ is available, and the ranking r with respect to unit k out of sample in some future period is of interest. The traditional way to address this problem is to model $p(Y|\mu) \propto f_\mu(Y)$, where $\mu = (\mu_1, ..., \mu_K)$ is a set of unit specified means, fit the model (probably in a Bayesian way), sample from the predictive posterior distribution, and then obtain the ranking based on μ_i naturally. The problem in this traditional way is, however, the ranking information in the data does not play a role in the likelihood directly, and hence the predictive posterior distribution of ranking could be 'blurred' during the Bayesian inference. That is, the information about the ranks is inducted through the likelihood $f_\mu(Y)$ and hence potentially a 'weak signal'. To let the ranking itself have a more significant influence in the model is what we want to address with the REL method.

In practice, the goal to rank K objects with no ties. The outcome of the procedure is a set of ranking which is defined as a permutation of the K rank indices. Each ranking has an associated ordering $\omega = (\omega_1, \omega_2, ..., \omega_K)$, which is defined as a permutation of the K item indices. In other words, it means that the item i is put in position ω_i. Consider the ranking of the observations $R = \{R_t, t = 1, 2, ..., T\}$ where R_t is the ranking of observation $y_t = \{y_{kt}, i = 1, 2, ..., K\}$. We blend this structure of ranking in the likelihood by the following definition with a function $g_\theta(R)$ naturally.

Definition: Assume

$$p(y_t|\mu, \theta) \propto f_\mu(y_t) g_\theta(r_t), \text{and hence } p(Y|\mu, \theta) \propto \Pi_t f_\mu(y_t) g_\theta(r_t) = f_\mu(Y) g_\theta(R),$$

where $f_\mu(Y) = \Pi_t f_\mu(y_t)$ and $g_\theta(R) = \Pi_t g_\theta(r_t)$.

Defining $c(\mu, \theta) = \int_z f_\mu(z) g_\theta(r(z)) dz$, where $z = \{z_k, k = 1, 2, ..., K\}$ is the observation in one time period and $r(z)$ is the corresponding ranking, then the Rank Enhanced Likelihood (REL) is

$$L(\mu, \theta|Y) \propto p(Y|\mu, \theta) = \frac{f_\mu(Y) g_\theta(R)}{c(\mu, \theta)^T}. \tag{4.1}$$

The definition implies that, the distribution of Y is multiplied by a factor $g_\theta(R)$ according to the rankings R, which allows the ranking to be directly considered in the likelihood and drive the inference of the model, rather than just being considered as the consequence of Y.

It is worth noting that, the Rank Enhanced distribution of Y is controlled by both μ and θ, where θ is the parameter which controls the weights of each rank directly. As a simple example, suppose $f_\mu(Y)$ is label-invariant for μ_k, which implies the same weighte on each and every rank. But with the $g_\theta(R)$, the distribution of Y's is reweighed according to the corresponding ranking R. With that said, it is important to note that the $g_\theta(R)$ is not the distribution of R. Rather the distribution of a given r is computed by

$p(r|\mu, \theta) = \int_{r(z)=r} \frac{f_\mu(z) g_\theta(r(z))}{c(\mu, \theta)} dy = \frac{g_\theta(r)}{c(\mu, \theta)} \int_{r(z)=r} f_\mu(z) dz.$

Another property is that, the conditional probability of an observation z given the ranking is

$p(z|\mu, \theta, r) = p(z|\mu, \theta)/p(r(z)|\mu, \theta) = f_\mu(z)/\int_{r(z)=r} f_\mu(z)dz,$

which means if the space of the observation is split according to different rankings, then within each slice of the space, the distribution of z is the same as $f_\mu(z)$.

4.2.2. Posterior Distribution

Supposing $\pi_\mu(\mu)$, $\pi_\theta(\theta)$ are the prior distributions of μ, θ respectively, then the posterior distribution corresponding to the REL is

$$p_{REL}(\mu, \theta|Y) \propto \frac{[\pi_\mu(\mu)\Pi_t f_\mu(y_t)][\pi_\theta(\theta)\Pi_t g_\theta(r_t)]}{(c(\mu, \theta))^T}. \tag{4.2}$$

Note that in equation (4.2), $\pi_\mu(\mu)\Pi_t f_\mu(y_t) \propto p(\mu|Y)$ and $\pi_\theta(\theta)\Pi_t g_\theta(r_t) \propto p(\theta|R)$, which suggests that in problems where simulation sampling from the initial posterior $p(\mu|Y)$ and $p(\theta|R)$ is available, importance sampling adjustments based on $w(\theta) = \frac{1}{(c(\mu,\theta))^T}$ can be used to compute $p_{REL}(\mu, \theta|Y)$ (Owen and Zhou 2000). Such simulation sampling can often be accomplished with MCMC algorithms, (Robert and Casella, 2004). Assuming for the moment that the value of $w(\theta)$ can be computed for any given μ and θ, the following importance sampling methods may be useful.

4.2.3. Obtain Samples and Estimates from Posterior Distribution

Suppose $\{\theta_1, \ldots, \theta_K\}$ is a simulated sequence of θ values that is converging in distribution to $p(\theta|Y)$ and $\{\mu_1, \ldots, \mu_K\}$ is a simulated sequence of μ values that is converging in distribution to $p(\mu|R)$, and let $\{w_1, \ldots, w_K\}$, where $w_k \equiv w(\mu_k, \theta_k)$, be the associated importance weights. Then, to compute posterior quantities of interest such as the posterior expectation of a function of the parameters $H(\theta, \mu)$,

$$E_{REL}H(\theta,\mu) \equiv \int_\theta H(\theta,\mu)p_{REL}(\mu,\theta|Y)d\mu d\theta, \tag{4.3}$$

the weighted sum

$$\frac{\Sigma_k H(\mu_k,\theta_k)w_k}{\Sigma_k w_k} \tag{4.4}$$

will be a consistent estimator of $E_{REL}H(\theta,\mu)$.

Going further, one can use the idea of sample importance resampling (SIR) to obtain a sample from $p_{REL}(\theta|Y)$ (Rubin et al. 1988, Smith and Gelfand 1992). Such a sample $\{\theta_1^*, \ldots, \theta_J^*\}$, can be obtained by sampling from $\{\theta_1, \ldots, \theta_K\}$ with replacement according to probabilities proportional to $\{w_1, \ldots, w_K\}$, and so for the parameter μ's. An attractive feature of such a resample is that it can be used to obtain a sample of predictions $\{Y_1^*, \ldots, Y_J^*\}$, where $Y_j^* \sim p(y|\mu_j^*, \theta_j^*)$. Note that $\{Y_1^*, \ldots, Y_J^*\}$ is effectively a sample from the REL predictive distribution

$$p_{REL}(Y^*|Y) = \int_\theta p(Y^*|\mu,\theta)p_{REL}(\mu,\theta|Y)d\mu d\theta. \tag{4.5}$$

The simulation sampling above facilities posterior inference about any function T of a prediction Y^* and/or parameters that captures an aspect of interest. Once $\{\theta_1^*, \ldots, \theta_J^*\}$, $\{\mu_1^*, \ldots, \mu_J^*\}$ and $\{Y_1^*, \ldots, Y_J^*\}$ have been obtained, simple substitution yields

$\{T(Y_1^*, \mu_1^*, \theta_1^*), \ldots, T(Y_J^*, \mu_J^*, \theta_J^*)\}$, a posterior-predictive sample of $T = T(Y^*, \mu, \theta)$ which can be used for further inference. In this chapter, we particularly focus on the prediction Y^* and the induced ranks $r(Y^*)$.

Lastly, we address the issue of computing $w(\mu, \theta)$ for any given pair of μ and θ, which is necessary for the implementation of the procedures above. For each pair of μ, θ, $c(\mu,\theta) = \int_z g_\theta(r(z))f_\mu(z)dz$. Accordingly, sampling z's from $f_\mu(z)$ and calculating $g_\theta(r(z))$, the mean of the sampled $g_\theta(r(z))$ is a consistent estimation of $c(\mu, \theta)$.

To perform the calculations in the simulations and applications in this chapter, we have used instances of the above methods. The same as the IRP method in Chapter 3, the sampling method could be improved for better efficiency.

4.3. Simulation

4.3.1. A Simple Example: $g_\theta(r)$ is a Multinomial Distribution Density Function

Suppose $K = 2$ items needs to be ranked, and the i.i.d. observations

$$p(y_t = (y_{1t}, y_{2t})|\mu, \theta) \propto f_\mu(y_t) g_\theta(r_t), \tag{4.6}$$

where $f_\mu(y_t)$ is the bivariate normal density function with mean $(0, 0)$ and identity covariance matrix, and $g_{\theta=(0.3,0.7)}(r)$ is the multinomial distribution density such $g_\theta(r_{y_{1t}>y_{2t}}) = 0.3$ and $g_\theta(r_{y_{1t}<y_{2t}}) = 0.7$.

Generating 100 datasets, observing the datasets are analyzed with both the standard Bayesian analysis and the REL analysis with the posterior sampling approaches proposed. We consider the posterior predictive probability of each ranking as the measurement of the goodness of the methods. From the standard Bayesian analysis where the effect of rankings is not considered in the likelihood, the posterior mean of μ is $(-0.266, 0.281)$ respectively for the 2 items and the posterior predictive probability of each ranking is $(0.354, 0.646)$ which is obviously 'flatter' than what it should be $(0.3, 0.7)$. It implies that skewing μ is not enough to recover the effect of the ranking part $g_\theta(r_t)$. From the posterior distribution fitted with REL, the posterior predictive probability of each ranking is $(0.276, 0.724)$, which skews the the estimation towards the more tilted ranking distribution.

4.3.2. A Hierarchical Linear Regression Simulation

4.3.2.1 Simulation Setup for $f_\mu(y)$

In this section, we conduct a simulation to examine the performance of the REL method with the example specified in Section 4.1. To conduct a simulation where the parameter values are 'realistic', the ERIM dataset, a 'classic and general' marketing scanner dataset, is studied with the method specified in Rossi et al. (2005) and the estimated parameters are used as the true parameters to simulate the data.

The ERIM dataset contains the grocery store scanner data collected by the now-defunct ERIM division of A.C. Nielsen on panels of households in two mid-sized Midwestern cities from 1985 through 1988, which records the sales of eight categories, including brownies, dinners, ketchup, margarine, peanut butter, sugar, toilet tissues and canned tuna. As an example, we study the sales of margarine. Thirty brands were records, and we chose to study the sales of the five brands with the largest sales volume. The weekly sales volumes are calculated, and composes the y_{it} sales matrix, where $i = 1, ..., 5$ denotes the brands and $t = 1, ..., 123$ denotes the index of weeks.

The dataset is studied with a Bayesian hierarchical linear regression model. The Bayesian hierarchical model has been successfully applied, for example, in gene selection (Bae and Mallick 2004), pricing strategies (Montgomery 1997), assortment choice (Bradlow and Rao 2000), products sales research (Buckley 1988), product consumption research (Yildiz Tiryaki and Akbay 2010), classification of financial policy (Dubas et al. 2010). See Rossi et al. (2005), Chapter 5 for an excellent overview.

Specifically, letting y_{it} denote the sales volumes of brand i at time t, $i = 1, \ldots, n = 5$, $t = 1, \ldots, T_1 = 123$, the distribution under the Bayesian hierarchical linear regression model is given by,

$$y_{it} = X_{it}\beta_i + \epsilon_{it}, \tag{4.7}$$

where X_{it} is a $J \times p$ design matrix, and $\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T$ is a brand-level parameter vector of length p where β_{i1} is the brand-specific intercept.

We further suppose that the brand-level coefficients in β_i follows the random effects formulation

$$\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T \sim Normal(\bar{\beta}, V_\beta), \tag{4.8}$$

$\bar{\beta} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)^T$ is the coefficient intercept vector, and

$V_\beta = diag(\sigma_1, \ldots, \sigma_p) \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1p} \\ \rho_{12} & 1 & \ldots & \rho_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{1p} & \rho_{2p} & \ldots & 1 \end{pmatrix} diag(\sigma_1, \ldots, \sigma_p)$ is the coefficient covariance matrix.

Letting $\mu = \{\bar{\beta}, V_\beta\}$ and $Y = \{y_{it}\}$, the model in equation (4.7) composes the $f_\mu(Y)$ part of the REL model. In this study, as an example we consider three major covariates, including the sales prices, an indicator whether a coupon was available, and an indicator whether an advertisement was conducted. These covariates are centered and standardized. Combined with the indicator for the intercept it composes the covariates matrix X_{it}. The model fit is conducted through MCMC sampling in R (R Development Core Team 2011). The estimated $\bar{\beta}$ and V_β serves as a set of true parameters to simulate the datasets used for out REL analyses.

4.3.2.2 Settings for $g_\theta(r)$

In situations where there are larger number of objects to rank, the multinomial model is not suitable for $p_\theta(r)$ since the number of possible rankings of K brands is $K!$, which increases fast with K. Besides this multivariate model, popular ranking models include a Plackett-Luce ranking model ((Plackett, 1975) and (Luce, 1959)) and Mallows Model ((Mallows, 1957), (Spearman, 1904) and (Kendall, 1938)). The Plackett-Luce (P-L) model has been applied in problems including modeling potential demand for electric cars ((Beggs et al., 1981)), modeling dietary preferences in cows ((Nombekela et al., 1994)), document ranking ((Cao et al., 2007)), etc.. In this simulation we apply the P-L model and use the

Bayesian method and prior distribution proposed in Guiver and Snelson (2009) to infer the parameters of the P-L model when there exists a larger number of customers.

We summarize the P-L model briefly as follows. Consider an experiment to rank K items with no ties. The outcome of the experiment is a set of rankings which is defined as a permutation of the K rank indices. Each ranking has an associated ordering $\omega = (\omega_1, \omega_2, ..., \omega_K)$, which is defined as a permutation of the K item indices. In other words, it means that the item i is put on position ω_i. The P-L model is a distribution over rankings which is parameterized by a vector of 'weights' $w = (w_1, w_2, ..., w_K)$ where $w_k > 0$ is associated with item k. The model is best described in term of the associated ordering ω:

$$PL(\omega|w) = \Pi_{k=1,2,...,K} f_k(w) \tag{4.9}$$
$$\text{where } f_k(w) = \frac{w_{\omega_k}}{w_{\omega_k} + w_{\omega_{k+1}} + ... + w_{\omega_K}}.$$

The P-L model can be intuitively interpreted as an urn model as in Silverberg (1980), where we consider a multi-stage experiment to draw balls from a urn of colorful balls. The number of balls of each color are in proportion to its 'weight' w_i. At the first stage a ball ω_1 is drawn from the urn, and the probability of this selection is $f_1(w_{\omega_1})$. At the k^{th} stage, another ball is drawn. If its color is drawn previously, put it back. Keep drawing balls until a new color ω_k is selected. Then the probability of this second selection is $f_k(w_{\omega_k})$. Continue until a ball of each color has been drawn. Obviously, equation (4.9) represents the probability of this color sequence.

Letting θ be the weights $w = (w_1, w_2, ..., w_K)$, we set $g_\theta(r_t) = PL(r_t|\theta = (w_1, w_2, ..., w_K))$. In the simulation, a variety of θ's are assumed to investigate the performance of REL.

4.3.2.3 Simulation Results

With the assumed $\bar{\beta}$, V_β and θ, one thousand sets of β_i and datasets y_{it} are generated according to model (4.1) where $f_\mu(y)$ is given by (4.7). Each generated dataset is analyzed with both the standard Bayesian approach and the REL approach, and the performance of both methods are compared.

For the standard Bayesian analysis, as a benchmark, the commonly used diffuse hyper-prior distributions are utilized as in Rossi et al. (2005).

$V_\beta \sim Inverse - Wishart(\nu, W)$, where $\nu = 5 + 3 = 8$ and $W = \nu I_5$;

$(\bar{\beta})|V_\beta \sim Normal(0, V_\beta \otimes 100 I_5)$.

We chose to measure the model performance by the MAE (Mean Absolute Error) of the out-of-sample predicted sales volume of the brands and their ranks. To illustrate the generalizability of the method, θ, the weights of the P-L model in $g_\theta(r)$ to range from flat $\theta \propto (1, 1, 1, 1, 1)$ to very skewed, for example $\theta \propto (1, 4, 9, 16, 25)$, and the data is simulated with the corresponding θ from the rank enhanced likelihood.

The percentage decrease of the MAE of the out-of-sample predicted sales volumes and ranks of the brands via REL compared to the standard Bayesian inference with a selection of θ is summarized as in Table 7. The improvement obtained from the REL approach is significant for both predicted sales volume and brand ranks.

Table 7: Percentage Decrease of the MAE of Predicted Sales Volume and Ranks

θ	Predicted Sales Volume	Predicted Ranks
$(1, 2, 3, 4, 5)$	14.59%	11.09%
$(1, 4, 9, 16, 25)$	17.08%	10.06%
$(5, 4, 3, 2, 1)$	4.94%	6.16%
$(25, 16, 9, 4, 1)$	5.29%	8.30%

4.4. ERIM dataset study

The ERIM dataset is studied with both the standard Bayesian method and the REL method. We measure model performance by MAE (Mean Absolute Error) of the out-of-sample predicted sales volume of the brands and their ranks. The percentage decrease of the MAE of the out-of-sample predicted sales volumes and ranks of the brands via REL compared to the standard Bayesian inference is 5.11% and 14.92%. If one incorporates a correlation structure $(\epsilon_{1t}, ...\epsilon_{5t}) \sim N(0, \Sigma_e)$, the percentage decrease of the MAE of the out-of-sample predicted sales volumes and ranks of the brands via REL compared to the standard Bayesian inference is 3.44% and 1.07%.

4.5. Discussion

The REL is an approach to incorporate the ranking structure in the likelihood, which enables the ranking to play a key role in the inference, rather than as a consequence of the predicted volume. The initial likelihood, aspect of inference interest and external information are not restricted and hence the method is flexible and generalizable. Our simulations demonstrate that REL can improve the predicted volume and ranks.

Another advantage of the REL approach is that it is easier to incorporate external information about the ranks by constructing an informative prior for θ. While the external information about the ranks may not be readily translatable to the prior of parameters μ in the initial model $f_\mu(y)$, it may be straightforward to be incorporated in an informative prior for θ.

The method is readily generalizable to the inference of other aspects of interest besides ranking, by replacing $g_\theta(r)$ with $g_\theta(t)$ where t is a function of the observation and/or parameters, for example partial ordering and classification. The inference and sampling approach remains the same, and hence the REL is a unified procedure to integrate an aspect of interest in the likelihood. Again, external information about an aspect of interest may be incorporated by constructing an informative prior for θ.

For computation, a simpler and more precise algorithm to sample from the posterior distribution may help improve the efficiency of the algorithm. In this study, we utilize importance sampling, which might be 'expensive' if the computation of the weights is not straightforward. The proposed distribution can certainly be improved for higher efficiency as mentioned in Section 3.2.1, and other sampling techniques may be a subject of future research. We believe this is a good first step, for those looking to incorporate rank (or other aspect) information directly.

CHAPTER 5 : Conclusion and Discussion

The major contribution of my dissertation consists of two methodologies: the IRP approach that incorporates external information in a unified process, and the REL approach that enables ranking information in a dataset to play a direct, key role in constructing models and that is readily generalizable. Both of the methods are obtained naturally during MCMC procedures which sample from the posterior distribution.

The IRP is a unified procedure to incorporate external information in a Bayesian inference structure. The likelihood, aspect of inference interest and external information are not restricted and hence the method is flexible and generalizable. Our simulations demonstrate that incorporating unbiased information through IRP can improve inference about a related aspect of interest. However, IRP may introduce bias in the inference of other parameters and hence the global fit, as explained in Section 3.2.1. The informative prior approach could involve a trade-off between global fit and a specific business problem, and the choice of the researcher/manager depends on the objective of the specific study.

There are two major questions remaining for the IRP method.

First, to what extent does the external information affect the estimation of the specific aspect of interest? The IRP utilizes information through the prior distribution, which could be restrictive when there is a large dataset, since the strong signal in the data may overwhelm the effect of external information. While this is a desired property of Bayesian analyses, this property restricts the potential effect of the external information. This problem is partly addressed by the rank enhanced likelihood method, which incorporates the aspect of interest directly and lets it play a key role in the inference.

Second, how to obtain and process the external information is not yet fully addressed in this study. In the application in Section 3.3, we obtain the external information in two ways: (1) from previous literature and (2) using part of the dataset. A certain degree of uncertainty is added to the obtained external information, since the datasets studied in the previous literature are not exactly the same as the dataset under research, and also the information obtained from part of the dataset may not precisely reflect the truth. A more rigorous external information process is important to fully develop; but for now, computing posterior distributions under a range of uncertainty levels is recommended.

With the questions said, we believe the IRP method is a good first step, for those looking to incorporate prior information on scales that are readily available, and are likely to exist in practice; for example the ways managers think as opposed to thinking on a parameter scale in which most priors are constructed.

The REL is an approach to incorporate the ranking structure in the likelihood, which enables the ranking to play a key role in the inference, rather than just see it as the consequence of the predicted volume. The initial likelihood, aspect of inference interest and external information are not restricted and hence the method is flexible and generalizable. Our simulations demonstrate that REL can improve the predicted volume and ranks. Another advantage of the REL approach is that it is easier to incorporate external information about the ranks by constructing an informative prior for θ. While the external information about the ranks may not be readily translatable to the prior of parameters, μ, in the initial model $f_\mu(y)$, it is straightforward to constructing an informative prior for θ to reflect the external information.

Both the IRP and rank enhanced likelihood methods requires adjustment to the posterior distribution, which is currently realized by importance sampling, which might be 'expensive' if the computation of the weights is not straightforward. A simpler and more precise algorithm to sample from the posterior distribution may help improve the efficiency of the algorithm. The proposed distribution can certainly be improved for higher efficiency as mentioned in Section 3.2.1, and other sampling techniques may be a subject of future research.

APPENDIX

A.1. Appendices For Chapter 2

A.1.1. Priors

To allow for posterior inferences to be mainly determined by the data, but to allow for shrinkage given the sparse data, we use proper, but diffuse priors, recognizing that the estimated parameter are on the logit scale. The priors were set as follows:

1. $\mu \sim N_4(0, 5I)$, where I is the identity matrix
2. $\Sigma_\mu \sim \text{Wishart}^{-1}(I,\ df = 9)$
3. $\Sigma_e \sim \text{Wishart}^{-1}(I,\ df = 10)$
4. $\beta_k) \sim N_6(0,\ 5I)$
5. $p_{active} \sim \text{Beta}(1,\ 1)$

Sensitivity analyses to these exact values, within a given range, indicated that the results were fairly robust to changes in the prior specification.

A.1.2. Simulation Study for Data Augmentation Procedure

To illustrate the identification of the model when one platform is observed in aggregate, and the effectiveness of the data augmentation method, we conducted a synthetic data parameter recovery study.

For our synthetic data set, setting $N = 200$, $K = 3$, $J = 30$, we generated observations y_{ikt} using the following true parameter values:

$$\mu = (-.5, -.3, -.8)$$

$$\sigma_\mu = \begin{pmatrix} .5 & .2 \\ .2 & .5 \end{pmatrix}$$

$$\sigma_e = \begin{pmatrix} .7 & .35 & .28 \\ .35 & .7 & .21 \\ .28 & .21 & .7 \end{pmatrix}$$

Since our focus was on understanding whether the covariance structure could be recovered, we did not use covariates in generating the synthetic data.

After generating individual-level y_{ikt} for all platforms, we collapsed the $k = 3$ platform into aggregated total usage $Y_{3t} = \Sigma_i y_{i3t}$. Assuming only the y_{ikt} where $k = 1, 2$ and Y_{i3t} are known to researchers, the synthetic dataset was fitted with the Bayesian data augmentation method described in the text.

Table 8 reports the estimated posteriors for the synthetic data and we find that the true parameters fall well within the posterior of the parameters. Notably, we are able to recover the covariance between the aggregate platform and the other platforms ($\Sigma_e[1, 3]$ and $\Sigma_e[2, 3]$), and the posteriors are only slightly wider than for the covariance between the two channels observed at the individual-level ($\Sigma_e[1, 2]$). We also found through other simulation studies that if we do not restrict the third column and row of Σ_μ to be 0, then this column and row is not identified by the data.

A.1.3. Procedure for Posterior Predictive Checks

We describe our procedure to generate tracking plots and posterior predictive outcomes. We begin with the posterior samples, obtained from the MCMC sampler, as described in the text. Following this, as is standard, we randomly select draws from our MCMC output and then simulate datasets from the model conditional on the drawn values of the parameters, as follows:

1. We randomly sample one draw from the MCMC output.

Table 8: Estimated Model Parameters: Synthetic Data

Parameters	True	Posterior Mean	2.5% -tile	97.5% -tile
$\mu[1]$	-.500	-.696	-.978	-.435
$\mu[2]$	-.300	-.120	-.379	.137
$\mu[3]$	-.800	-.695	-.999	-.393
$\Sigma_\mu[1,1]$	.500	.458	.327	.630
$\Sigma_\mu[1,2]$	.200	.163	.0585	.282
$\Sigma_\mu[2,2]$	.500	.605	.390	.871
$\Sigma_e[1,1]$	.700	.535	.306	.917
$\Sigma_e[1,2]$	.350	.215	.041	.462
$\Sigma_e[1,3]$	.280	.251	.030	.559
$\Sigma_e[2,2]$	.700	.433	.230	.772
$\Sigma_e[2,3]$	.210	.196	-.0099	.478
$\Sigma_e[3,3]$	.700	.759	.440	1.285

2. With the set of sampled parameters, we generate a pseudo dataset of observations of users Y^* according to the model in Section 3, and of the same size as the observed data set.

3. For each set of generated observations of users, we calculate statistic $T^*|Y^*$.

4. Repeat step 2-4 100 times, and compute statistics T^*_{iter} for $iter = 1, 2, ..., 100$.

5. We compare the generated statistics T^*_{iter} for $iter = 1, 2, ..., 100$ with the corresponding statistic calculated from the true dataset $T|Y$ to obtain the appropriate Bayesian p-values.

A.1.4. Other Posterior Predictive Checks

Our other posterior predictive checks focus on two key questions important to media planning: 1) Do users tend to "jump in" or "drop out" during the tournament? and 2) How many heavy users are there and does this vary by platform?

Users "jumping in" and "dropping out". Media planners are commonly interested in the "big game effect," i.e., how many users ignore the tournament until the later elimination games and then "jump in" for the "big games." In Table 9, we report the actual and predicted number of users who watched *only* on the days of the quarter-finals, semi-finals and final games or *only* on the days of the semi-final and final games. Note that there are no features of our model that directly relate to these measures, rather the predicted number of users who watch only during the final stage is a complex function of the observed covariates (x_{kt}) for those days and the heterogeneity across users. While there is a slight underprediction for both of these statistics, we find that the model fits the data remarkably well given that there are no parameters of the model that directly relate to the tournament stages.

Similarly, since our data is a sample of US viewers, there is also a "big game effect" on days when the US team plays, and there could be a number of users who view *only* on days when the US is playing. When we compare the predicted and actual number of users who only watch on the days when the US team plays, we find that we slightly underpredict this effect, suggesting that the US team effect may be slightly underestimated.

Finally, a critical issue for media planners is whether there is a point during the tournament when users lose interest and drop out (i.e., stop watching entirely). In Table 9 we report the number of users who stop watching after the group stage when one-half of the original thirty-two teams are eliminated, and the number of users who drop out after the US team is eliminated. Again, this type of drop-out is not naturally related to a single model parameter. We note that a substantial number of viewers do drop out at this stage and we slightly under predict this. So, there may be some specific effects of the tournament stages that the model does not capture, but overall, we find a reasonably good model fit.

Heavy users by platform. Media planners are also interested in identifying heavy users. Heavy users, those who access a platform more frequently, are more likely to see an advertisement placed on a particular platform repeatedly, which makes them more likely to take action on the advertised product or service (Rossiter and Danaher (1998)). Thus, media planners want to know how many heavy users there are and which platforms they visit most often. Table 10 reports the actual and posterior mean for the percentage of

Table 9: Posterior Predictive Checks Related to User Drop-out and Jump-in

	Actual	Predicted	Posterior Quantile
Watch only during the quarter-final, semi-final and final games?	2.70%	1.93%	.93
Watch only during the semi-final and final games?	1.75%	1.11%	.96
Watch only on days when the US is playing?	1.90%	1.52%	.83
Stop watching after the group stage?	9.90%	8.82%	.86
Stop watching after the US drops out?	1.80%	9.52%	.93

users who view content on more than 20% and 50% of the days for each platform. We find that the model is able to fit these metrics well for the lower threshold, suggesting that the model accurately captures the heterogeneity across users. For the 50% threshold, which represents the extreme tail of very heavy users, we find that the model makes a reasonable prediction for the ESPN.com channel, but does much worse at predicting the two less-used digital channels. This may suggest that the tail of the Gaussian distribution is somewhat "too thick." However, overall, these results suggest that the specification we have predicts differences among users reasonably well.

Table 10: Posterior Predictive Checks Related to Heavy Versus Light Users

	Actual	Predicted	Posterior Quantile
How many people watch ESPN.com on more than 20% of days?	6.9%	6.6%	.59
How many people watch streaming ESPN3 on more than 20% of days?	6.4%	5.8%	.77
How many people watch ESPN Mobile on more than 20% of days?	2.2%	1.1%	.33
How many people watch ESPN.com on more than 50% of days?	1.8%	2.0%	.25
How many people watch streaming ESPN3 on more than 50% of days?	.7%	1.4%	.02
How many people watch ESPN Mobile on more than 50% of days?	.1%	.17%	.18

To illustrate the limitations of the model, we have reported here several posterior predictive checks that were less successful. The posterior prediction for how many people watch only when the US team is playing is lower than the actual data. We also find limitations in predicting the amount of drop-out at the end of the tournament and predicting the number of heavy users for the lesser-used platforms suggesting that the Gaussian distribution may not be fitting the tails of the user-distribution perfectly. Despite these (minor) deficiencies, we feel the posterior predictive checks demonstrate the overall suitability of the model to the data.

BIBLIOGRAPHY

G. Allenby and P. Rossi. Quality perceptions and asymmetric switching between brands. *Marketing Science*, 10:185–204, 1991.

A. Ansari, C. Mela, and S. Neslin. Customer channel migration. *Journal of Marketing Research*, 45(1):60–76, 2008.

E. Arjas and D. Gasbarra. Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association*, pages 1101–1109, 1996.

K. Bae and B. Mallick. Gene selection using a two-level hierarchical bayesian model. *Bioinformatics*, 20(18):3423–3430, 2004.

S. Beggs, S. Cardell, and J. Hausman. Assessing the potential demand for electric cars. *Journal of econometrics*, 17(1):1–19, 1981.

P. Boatwright, R. McCulloch, and P. Rossi. Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *Journal of the American Statistical Association*, 94(448):1063–1073, 1999.

E. Bradlow and V. Rao. A hierarchical bayes model for assortment choice. *Journal of Marketing Research*, pages 259–268, 2000.

E. Bradlow and A. Zaslavsky. Case influence analysis in bayesian inference. *Journal of Computational and Graphical Statistics*, 6:314–331, 1997.

M. Braun and W. Moe. Online advertising campaigns: Modeling the effects of multiple ad creatives. 2011.

R. Briggs. Abolish clickthrough now! *accessed October*, 16:2005, 2001.

P. Buckley. Nested multinomial logit analysis of scanner data for a hierarchical choice model. *Journal of Business Research*, 17:133–154, 1988.

Z. Cao, T. Qin, T. Liu, M. Tsai, and H. Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on Machine learning*, pages 129–136. ACM, 2007.

M. Chen and J. Ibrahim. The relationship between the power prior and hierarchical models. *Bayesian Analysis*, 1(3):551–574, 2006.

Y. Chen and S. Yang. Estimating disaggregate models using aggregate data through augmentation of individual choice. *Journal of Marketing Research*, 44(4):613–621, 2007.

S. Chib and E. Greenberg. Analysis of multivariate probit models. *Biometrika*, 85(2): 347–361, 1998.

A. Crupi. Espn fields major sports research initiative. *AdWeek*, pages March 23, 2010, 2010.

P. Danaher. Modeling page views across multiple websites with an application to internet reach and frequency prediction. *Marketing Science*, 26(3):422–437, 2007.

P. Danaher and D. Mawhinney. Optimizing television program schedules using choice modeling. *Journal of Marketing Research*, pages 298–312, 2001.

P. Danaher, T. Dagger, and M. Smith. Enoch, glenn. *Presentation to the 2009 Joint Statistical Meetings, Washington, DC*, 2009.

W. DeSarbo, P. Ebbes, D. Fong, and C. Snow. Revisiting customer value analysis in a heterogeneous market. *Journal of Modelling in Management*, 5(1):8–24, 2010.

J. Dow and J. Endersby. Multinomial probit and multinomial logit: a comparison of choice models for voting research. *Electoral Studies*, 23(1):107–122, 2004.

X. Dreze and F. Hussherr. Internet advertising: Is anybody watching? *Journal of interactive marketing*, 17(4):8–23, 2003.

J. Dubas, B. Lee, and N. Mark. A multinomial logit approach to exchange rate policy classification with an application to growth. *Journal of International Money and Finance*, 29(7):1438–1462, 2010.

G. Franz. The future of multimedia research. *International journal of Market Research*, 42 (4):459–472, 2000.

A. Gelfand and A. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–533, 2006.

A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

A. Gelman, X. Meng, and H. Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6:733–759, 1996.

A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian data analysis*. Chapman & Hall/CRC, 2003.

A. Gelman, A. Jakulin, M. Pittau, and Y. Su. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4): 1360–1383, 2008.

G. Glonek and P. McCullagh. Multivariate logistic models. *Journal of the royal statistical society. Series B (Methodological)*, pages 533–546, 1995.

J. Guiver and E. Snelson. Bayesian inference for plackett-luce ranking models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 377–384. ACM, 2009.

J. Higgins and A. Whitehead. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, 15(24):2733–2749, 1996.

F. Hofstede, Y. Kim, and M. Wedel. Bayesian prediction in hybrid conjoint analysis. *Journal of Marketing Research*, 39(2):253–261, 2002.

J. Ibrahim and M. Chen. Prior distributions and bayesian computation for proportional hazards models. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 48–64, 1998.

J. Ibrahim and M. Chen. Power prior distributions for regression models. *Statistical Science*, pages 46–60, 2000.

K. C. W. Joo, Mingyu and Y. Zhu. Does television advertising influence online search? *Available at SSRN: http://ssrn.com/abstract=1720713*, 2011.

M. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

P. Lenk and B. Orme. The value of informative priors in bayesian inference with sparse data. *Journal of Marketing Research*, 46(6):832–845, 2009.

P. Lenk and A. Rao. New models from old: Forecasting product adoption by hierarchical bayes procedures. *Marketing Science*, 9:42–53, 1990.

P. Lenk, W. DeSarbo, P. Green, and M. Young. Hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science*, 15(2):173–191, 1996.

C. Lin, S. Venkataraman, and S. Jap. Media multiplexing behavior: Implications for targeting and media planning. *Working Paper*, 2010.

S. Lind and J. Kivisto-Rahnasto. Utilization of external accident information in companies' safety promotion-case: Finnish metal and transportation industry. *Safety Science*, 46(5): 802–814, 2008.

R. Little and D. Rubin. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, 37(3):218–220, 1983.

R. Luce. Individual choice behavior. 1959.

T. Lybbert, C. Barrett, J. McPeak, and W. Luseno. Bayesian herders: Updating of rainfall beliefs in response to external forecasts. *World Development*, 35(3):480–497, 2007.

S. MacEachern and M. Peruggia. Subsampling the gibbs sampler: Variance reduction. *Statistics & probability letters*, 47(1):91–98, 2000.

C. Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.

P. Manchanda, A. Ansari, and S. Gupta. The shopping basket: A model for multicategory purchase incidence decisions. *Marketing Science*, 18(2):95–114, 1999.

P. Manchanda, J. Dubé, K. Goh, and P. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1):98–108, 2006.

W. Moe and P. Fader. Dynamic conversion behavior at e-commerce sites. *Management Science*, pages 326–335, 2004.

A. Montgomery. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science*, 16(4):315–337, 1997.

A. Montgomery and P. Rossi. Estimating price elasticities with theory-based priors. *Journal of Marketing Research*, 36:413–423, 1999.

A. Musalem, E. Bradlow, and J. Raju. Who's got the coupon? estimating consumer preferences and coupon usage from aggregate information. *Journal of Marketing Research*, 45(6):715–730, 2008.

A. Musalem, E. Bradlow, and J. Raju. Bayesian estimation of random-coefficients choice models using aggregate data. *Journal of Applied Econometrics*, 24(3):490–516, 2009.

P. Naik and K. Peters. A hierarchical marketing communications model of online and offline media synergies. *Journal of Interactive Marketing*, 23(4):288–299, 2009.

P. Naik and K. Raman. Understanding the impact of synergy in multimedia communications. *Journal of Marketing Research*, pages 375–388, 2003.

S. Nombekela, M. Murphy, H. Gonyou, and J. Marden. Dietary preferences in early lactation cows as affected by primary tastes and some common feed flavors. *Journal of dairy science*, 77(8):2393–2399, 1994.

C. O'Donnell and T. Coelli. A bayesian approach to imposing curvature on distance functions. *Journal of Econometrics*, pages 493–523, 2005.

A. Owen and Y. Zhou. Safe and effective importance sampling. *Journal of the American Statistical Association*, 95:135–143, 2000.

G. Parker and M. Van Alstyne. Two-sided network effects: A theory of information product design. *Management Science*, 51(10):1494–1504, 2005.

M. Peruggia. On the variability of case-deletion importance sampling weights in the bayesian linear model. *Journal of the American Statistical Association*, 92:199–207, 1997.

J. Pilotta and D. Schultz. Simultaneous media experience and synesthesia. *JOURNAL OF ADVERTISING RESEARCH-NEW YORK-*, 45(1):19, 2005.

J. Pilotta, D. Schultz, G. Drenik, and P. Rist. Simultaneous media usage: A critical consumer orientation to media planning. *Journal of Consumer Behaviour*, 3(3):285–292, 2004.

R. Plackett. The analysis of permutations. *Applied Statistics*, pages 193–202, 1975.

D. Putler, K. Kalyanam, and J. Hodges. A bayesian approach for estimating target market potential with limited geodemographic information. *Journal of Marketing Research*, 33: 134–149, 1996.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, 2004.

P. Rossi and G. Allenby. A bayesian approach to estimating household parameters. *Journal of Marketing Research*, pages 171–182, 1993.

P. Rossi, G. Allenby, and R. McCulloch. *Bayesian statistics and marketing*. Wiley New York:, 2005.

J. Rossiter and P. Danaher. *Advanced media planning*. Springer, 1998.

D. Rubin et al. Using the sir algorithm to simulate posterior distributions. *Bayesian statistics*, 3:395–402, 1988.

Z. Sandor and M. Wedel. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, 28:430–444, 2001.

B. Sharp and A. Sharp. Loyalty programs and their impact on repeat-purchase loyalty patterns. *International Journal of Research in Marketing*, 14(5):473–486, 1997.

L. Sheiner and S. Beal. Some suggestions for measuring predictive performance. *Journal of Pharmacokinetics and Pharmacodynamics*, 9(4):503–512, 1981.

S. Shim and M. Mahoney. The elderly mail-order catalog user of fashion products:: A profile of the heavy purchaser. *Journal of Direct Marketing*, 6(1):49–58, 1992.

A. Smith and A. Gelfand. Bayesian statistics without tears: a sampling-resampling perspective. *American statistician*, pages 84–88, 1992.

Y. Song. Proof that online advertising works. *Atlas Institute, Seattle, WA*, 2001.

C. Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.

D. Speigelhalter, A. Thomas, and N. Best. Winbugs version 1.2 user manual. *Cambridge: Medical Research Council Biostatistics Unit*, 1999.

A. Sutton and K. Abrams. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10(4):277, 2001.

M. Tanner and W. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, 82(398):528–540, 1987.

T. Trikalinos, G. Salanti, E. Zintzaras, and J. Ioannidis. Meta-analysis methods. *Advances in genetics*, 60:311–334, 2008.

M. Wedel and R. Pieters. Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science*, 19:297–312, 2000.

G. Yildiz Tiryaki and C. Akbay. Consumers fluid milk consumption behaviors in turkey: an application of multinomial logit model. *Quality & Quantity*, 44(1):87–98, 2010.

E. Zheng, P. Fader, and B. Padmanabhan. From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. *Available at SSRN 1323587*, 2009.